

UNIVERSIDADE FEDERAL DO PARANÁ

JHONNY IVAIR DE LIMA MACIEL

INDICADORES PARA O AGRONEGÓCIO UTILIZANDO
FERRAMENTA WEB DE VISUALIZAÇÃO DA
INFORMAÇÃO

CURITIBA

2019

JHONNY IVAIR DE LIMA MACIEL

INDICADORES PARA O AGRONEGÓCIO UTILIZANDO FERRAMENTA WEB DE VISUALIZAÇÃO DA INFORMAÇÃO

Dissertação apresentada ao curso de Pós-Graduação em Ciência, Gestão e Tecnologia da Informação, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Ciência, Gestão e Tecnologia da Informação.

Orientador: Prof. Dr. Egon Walter Wildauer

CURITIBA

2019

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DE CIÊNCIAS SOCIAIS
APLICADAS – SIBI/UFPR COM DADOS FORNECIDOS PELO(A) AUTOR(A)
Bibliotecário: Eduardo Silveira – CRB 9/1921

Maciel, Jhonny Ivair de Lima

Indicadores para o agronegócio utilizando ferramenta web de
visualização da informação / Jhonny Ivair de Lima Maciel . - 2019.
240 p.

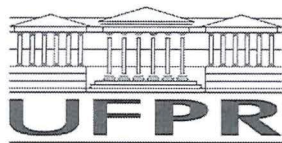
Dissertação (Mestrado) – Universidade Federal do Paraná.
Programa de Pós- Graduação em Ciência, Gestão e Tecnologia da
Informação, do Setor de Ciências Sociais Aplicadas.

Orientador: Egon Walter Wildauer.

Defesa: Curitiba, 2019.

1. Gestão da Informação. 2. Agronegócios. 3. Indicadores. I.
Universidade Federal do Paraná. Setor de Ciências Sociais Aplicadas.
Programa de Pós- Graduação em Ciência, Gestão e Tecnologia da
Informação. II. Wildauer, Egon Walter. III. Título.

CDD 658.4038




MINISTÉRIO DA EDUCAÇÃO
SETOR SETOR DE CIÊNCIAS SOCIAIS E APLICADAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO GESTÃO DA
INFORMAÇÃO - 40001016058P1

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GESTÃO DA INFORMAÇÃO da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **JHONNY IVAIR DE LIMA MACIEL** intitulada: **INDICADORES PARA O AGRONEGÓCIO UTILIZANDO FERRAMENTA WEB DE VISUALIZAÇÃO DA INFORMAÇÃO**, após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 28 de Março de 2019.


EGON WALTER WILDAUER
Presidente da Banca Examinadora (UFPR)


NORBERTO FERNANDO KUCHENBECKER
Avaliador Externo (F. ESTÁCIO)


JOSÉ SIMÃO DE PAULA PINTO
Avaliador Interno (UFPR)


CARLOS HENRIQUE KURETZKI
Avaliador Externo (UP)

*Este trabalho é dedicado à Ana Gomes Vieira Corrêa e à Polly,
que nos deixaram durante o desenvolvimento da pesquisa.*

AGRADECIMENTOS

Primeiramente à Deus, pela saúde que me concedeu, além de todas as oportunidades que me ofereceu através de sua imensa bondade.

A minha amada esposa Joicy, por todo o carinho, companheirismo e apoio diário que me deu neste período de forma incondicional, fazendo com que eu nunca pensasse em desistir.

Minha mãe, Ivana, que fez de tudo por mim desde o princípio. Devo tudo o que sou e o que conquistei a esta mulher, que me ensinou tudo que sei.

Ao meu orientador, Professor Egon, por toda a paciência e auxílio, fundamentais para a construção deste trabalho. Agradeço também aos demais professores do Programa de Pós-Graduação em Gestão da Informação por todo o conhecimento que transmitem em suas aulas, em especial ao Professor Simão e Professora Helena, sempre extremamente solícitos aos alunos.

Aos meus colegas de mestrado Felipe Fiori e Lucas, sempre dispostos a discutir melhorias, compartilhar conhecimentos e boas conversas.

Aos meus colegas de trabalho da equipe SIGA, sobretudo Mattheus, Nickolas e Lucas pelas críticas sempre construtivas, contribuições de ideias para melhora do resultado final e direcionamentos técnicos com a Linguagem R, da qual são extremamente capazes de propor soluções de qualquer tipo além de compartilharem todo este conhecimento.

Ao meu supervisor José Henrique, o primeiro incentivador de meu ingresso no mestrado, sempre confiando em minha capacidade, além de proporcionar o tempo e prioridade necessários para a construção desta pesquisa.

RESUMO

A cadeia produtiva do agronegócio no Brasil é responsável por entre 20 a 30% do Produto Interno Bruto desde a década de 90, e envolve desde a produção de insumos e maquinários para utilização na agricultura e pecuária até o consumidor final. A competitividade no setor faz com que a utilização da informação para suporte a decisão seja fator de vantagem entre os atores envolvidos com o agronegócio, podendo, ao mesmo tempo, servir também na tomada de decisão do poder público. Alguns órgãos brasileiros ligados ao agronegócio disponibilizam, com intuito de apoio a decisão, ferramentas de visualização de informação de seus dados. Estas ferramentas porém, deixam algumas lacunas, que ao serem exploradas possibilitam maior auxílio à usuários da informação em futuras decisões. Portais como a CONAB, EMBRAPA, IBGE e FAO disponibilizam dados brutos de produção agrícola, que se agregados em um único conjunto de dados torna-se uma robusta fonte de informação, decorrente da variedade de dados acerca de diferentes produtos em épocas distintas. A fonte de informações aliada ao uso de técnicas de visualização da informação disponibiliza ao usuário a possibilidade de filtragem e melhor compreensão dos dados, sempre com intuito de proporcionar maior suporte a decisão nos âmbitos público e privado. Além da questão da tomada de decisão nos setores público e privado, ferramentas que se utilizam de visualização da informação podem apontar fenômenos presentes em seus conjuntos de dados, que posteriormente pode ser estudados e explicados. É objetivo deste trabalho demonstrar o desenvolvimento de uma ferramenta web que disponibilize indicadores de produção do agronegócio, realizando funcionalidades complementares as disponibilizadas por ferramentas governamentais, como a utilização de dados de diferentes fontes, possibilidade de cruzamento de dados de diferentes produtos relacionados a pecuária e a agricultura do Brasil, utilizando também métodos estatísticos para descrição dos dados.

Palavras-chaves: Gestão da informação; Agronegócio; Indicador; Integração de dados.

ABSTRACT

The agribusiness production chain in Brazil accounts for between 20-30% of the gross domestic product since the 1990s, and involves everything from the production of inputs and machinery for use in agriculture and livestock to the final consumer. Competitiveness in the sector means that the use of information to support decision making is an advantage factor among the actors involved in agribusiness, while at the same time, it can also serve as a decision-maker for public power. Some Brazilian agencies related to agribusiness offer, in order to support the decision, tools to display information of their data. These tools, however, leave some gaps that, when explored, allow greater assistance to information users in future decisions. Portals such as CONAB, EMBRAPA, IBGE and FAO provide raw agricultural production data, which if aggregated into a single data set becomes a robust source of information, resulting from the variety of data about different products at different times. The source of information combined with the use of information visualization techniques provides the user with the possibility of filtering and better understanding of the data, always with the purpose of providing greater decision support in the public and private spheres. In addition to the issue of decision making in the public and private sectors, tools that use information visualization can point to phenomena present in their data sets, which can later be studied and explained. The objective of this work is to demonstrate the development of a web tool that provides agribusiness production indicators, performing complementary functionalities provided by governmental tools, such as the use of data from different sources, the possibility of crossing data from different products related to livestock and also using statistical methods to describe the data.

Key-words: Information Management; Agribusiness; Performance Indicator; Data Integration.

LISTA DE FIGURAS

Figura 1 – OPERAÇÃO GENÉRICA DA CADEIA PRODUTIVA AGROINDUSTRIAL	20
Figura 2 – DISTRIBUIÇÃO DE TRABALHOS QUE ABORDAM AGRONEGÓCIO E INDICADORES DESDE A DÉCADA DE 90	25
Figura 3 – ESTRUTURA DO TRABALHO	28
Figura 4 – ESTRUTURA DO REFERENCIAL TEÓRICO	30
Figura 5 – GRANDES CENTROS DE AGRICULTURA NO PERÍODO NEOLÍTICO	31
Figura 6 – POPULAÇÃO URBANA E RURAL DOS ANOS 60 AOS 2000	36
Figura 7 – CADEIA PRODUTIVA AGROINDUSTRIAL	38
Figura 8 – CADEIA PRODUTIVA AGROINDUSTRIAL	40
Figura 9 – ÁREAS DE CONHECIMENTO DA GESTÃO DA INFORMAÇÃO	43
Figura 10 – PROCESSO DE TRANSFORMAÇÃO DA INFORMAÇÃO	46
Figura 11 – MODELO DIKAR E RAKID	46
Figura 12 – CICLO DE VIDA DA INFORMAÇÃO	47
Figura 13 – CICLO DA UTILIZAÇÃO DE FONTES EXTERNAS DE DADOS	59
Figura 14 – ESTÁGIOS DOS DADOS NA ETL	60
Figura 15 – TRANSFORMAÇÃO DE DADOS POR <i>TIDY DATA</i>	62
Figura 16 – FORMATO DO OBJETO JSON DE TROCA DE DADOS	64
Figura 17 – PROCESSO DE TRANSFORMAÇÃO DE DADOS EM REPRESENTAÇÃO GRÁFICA	72
Figura 18 – EXEMPLO DE BOXPLOT	76
Figura 19 – EXEMPLO DE GRÁFICO DE BARRAS VERTICAL E HORIZONTAL	76
Figura 20 – EXEMPLO DE HISTOGRAMA	77
Figura 21 – EXEMPLO DE GRÁFICO DE DISPERSÃO	77
Figura 22 – EXEMPLO DE GRÁFICO DE LINHAS	78
Figura 23 – EXEMPLO DE GRÁFICO DE PIZZA	79
Figura 24 – EXEMPLO DE <i>DASHBOARD</i>	80
Figura 25 – SEQUÊNCIA LÓGICA DOS MATERIAIS E MÉTODOS UTILIZADOS NA DISSERTAÇÃO	81
Figura 26 – FLUXO DOS MÉTODOS DA PESQUISA	84
Figura 27 – FLUXOGRAMA MODELO DO ALGORITMO DE EXTRAÇÃO DOS DADOS	86
Figura 28 – <i>DASHBOARD</i> SHINYAPPS	91
Figura 29 – EXECUÇÃO DE APLICATIVOS SHINY NO RSTUDIO	92
Figura 30 – CONTROLE DE CONTAS SHINYAPPS LIGADAS AO RSTUDIO	93
Figura 31 – INSERÇÃO DO <i>TOKEN</i> DO SHINYAPPS NO RSTUDIO	94
Figura 32 – PUBLICAÇÃO DE APLICAÇÃO NO SHINYAPPS VIA RSTUDIO	94

Figura 33 – DADOS DISPONIBILIZADOS PELA CONAB	96
Figura 34 – SELEÇÃO DE LOCAL PARA EXPORTAÇÃO DE DADOS DA EMBRAPA	99
Figura 35 – DISPONIBILIZAÇÃO DE DADOS NO PORTAL SIDRA - IBGE	100
Figura 36 – CÓDIGO FONTE DO PORTAL DA CONAB	103
Figura 37 – FLUXOGRAMA DE EXTRAÇÃO DE DADOS CONAB	104
Figura 38 – SELEÇÃO DOS ESTADOS PARA EXTRAÇÃO DE DADOS DA EMBRAPA	105
Figura 39 – <i>DOWNLOAD</i> DO ARQUIVO COM DADOS DA EMBRAPA	106
Figura 40 – REDEFINIÇÃO DE PARÂMETROS PARA OBTENÇÃO DE DADOS DA EMBRAPA	106
Figura 41 – DISPONIBILIZAÇÃO DOS DADOS DA PLATAFORMA SIDRA - IBGE	107
Figura 42 – ENDEREÇO ELETRÔNICO PARA OBTENÇÃO DOS DADOS DO IBGE	107
Figura 43 – FLUXOGRAMA DE EXTRAÇÃO DE DADOS IBGE	108
Figura 44 – ENDEREÇO ELETRÔNICO DA FAO PARA ACESSO AO JSON COM DADOS	109
Figura 45 – FLUXOGRAMA DE EXTRAÇÃO DE DADOS FAO	110
Figura 46 – ESTRUTURA DA PLANILHA COM DADOS DA CONAB	112
Figura 47 – FLUXOGRAMA DE LIMPEZA DE DADOS CONAB	113
Figura 48 – ESTRUTURA DA PLANILHA COM DADOS DA EMBRAPA	114
Figura 49 – PROBLEMA DE CÁLCULO DA PRODUTIVIDADE NOS DADOS DA EMBRAPA	115
Figura 50 – FLUXOGRAMA DE LIMPEZA DE DADOS EMBRAPA	116
Figura 51 – ESTRUTURA DA PLANILHA COM DADOS DO IBGE	117
Figura 52 – FLUXOGRAMA DE LIMPEZA DE DADOS IBGE	119
Figura 53 – ESTRUTURA DO JSON DA FAO	121
Figura 54 – FLUXOGRAMA DE LIMPEZA DE DADOS FAO	121
Figura 55 – <i>DATA FRAME</i> COM DADOS DA FAO PÓS ITERAÇÃO	122
Figura 56 – SEPARAÇÃO DE ABAS DA FERRAMENTA	124
Figura 57 – EXEMPLO DE GRÁFICO GERADO PELA FERRAMENTA	125
Figura 58 – EXEMPLO DE TABELA ESTATÍSTICA GERADA PELA FERRAMENTA	126
Figura 59 – EXEMPLO DE <i>BOXPLOT</i> GERADO PELA FERRAMENTA	127
Figura 60 – FILTROS DISPONIBILIZADOS PELA FERRAMENTA	128
Figura 61 – VISÃO GERAL DA FERRAMENTA	129
Figura 62 – DISPONIBILIZAÇÃO DOS METADADOS DA FERRAMENTA	130
Figura 63 – ROTINA PARA COLETA E LIMPEZA DOS DADOS	130
Figura 64 – ROTINA PARA COLETA E LIMPEZA DOS DADOS	131
Figura 65 – AUMENTO DA ÁREA COLHIDA DE CANA-DE-AÇÚCAR EM MATO GROSSO DO SUL ENTRE 2003 E 2011	132

Figura 66 – VARIAÇÃO DE PRODUTIVIDADE DE CANA-DE-AÇÚCAR ENTRE 2010 E 2011 EM MATO GROSSO DO SUL	132
Figura 67 – VARIAÇÃO DE PRODUÇÃO DE CANA-DE-AÇÚCAR ENTRE 2001 E 2009 NO CENTRO-SUL DO BRASIL	133
Figura 68 – RESULTADOS DA FERRAMENTA PARA PRODUTIVIDADE DE SOJA NO BRASIL ENTRE 1964 E 2005	134
Figura 69 – AUMENTO NA PRODUÇÃO DE SOJA NO ESTADO DE MATO GROSSO A PARTIR DO ANO DE 1997	135
Figura 70 – ESTADOS PRODUTORES DE SOJA EM 2015 SEGUNDO A FERRAMENTA ATRAVÉS DE DADOS DO IBGE	137
Figura 71 – ESTADOS PRODUTORES DE SOJA EM 2015 SEGUNDO A FERRAMENTA ATRAVÉS DE DADOS DA CONAB	137
Figura 72 – ESTADOS PRODUTORES DE SOJA EM 2015 SEGUNDO A EMBRAPA UTILIZANDO A FERRAMENTA	138
Figura 73 – DADOS DE ÁREA PLANTADA DE SOJA DISPONIBILIZADOS PELA CONAB	139
Figura 74 – EVOLUÇÃO DA ÁREA PLANTA DE SOJA NO BRASIL ENTRE 1990 E 2005	140

LISTA DE TABELAS

Tabela 1 – PESQUISA DOS TERMOS "GESTÃO DA INFORMAÇÃO" E "AGRO-NEGÓCIO"	24
Tabela 2 – PESQUISA DOS TERMOS "AGRONEGÓCIO" E "INDICADORES" .	25
Tabela 3 – PESQUISA DOS TERMOS "AGRONEGÓCIO" E "VISUALIZAÇÃO DA INFORMAÇÃO"	26
Tabela 4 – RESULTADOS DA PESQUISA SOBRE AGRONEGÓCIO, VISUALIZAÇÃO DA INFORMAÇÃO E INDICADORES	26
Tabela 5 – COMPARAÇÃO DE RESULTADOS DA FERRAMENTA E DO TRABALHO DE CONTINI ET AL., 2006 - PRODUTIVIDADE DE SOJA EM TONELADAS POR HECTARE NO BRASIL	134
Tabela 6 – CONTRIBUIÇÃO POR REGIÃO NA PRODUÇÃO DE SOJA DO BRASIL EM 2015	138
Tabela 7 – COMPARAÇÃO DE RESULTADOS DE PRODUÇÃO INDICADOS NO TRABALHO DE FERREIRA 2012 E DA FERRAMENTA - DADOS DO BRASIL EM 2006	140
Tabela 8 – COMPARAÇÃO DE RESULTADOS DE PRODUÇÃO (EM MIL TONELADAS) INDICADOS NO TRABALHO DE GASQUES ET. AL. 2004 E DA FERRAMENTA - DADOS DO BRASIL EM 2003	141
Tabela 9 – COMPARAÇÃO DE RESULTADOS DE ÁREA COLHIDA (EM MIL HECTARES) INDICADOS NO TRABALHO DE GASQUES ET. AL. 2004 E DA FERRAMENTA - DADOS DO BRASIL EM 2003	142
Tabela 10 – COMPARAÇÃO DE RESULTADOS DE PRODUTIVIDADE (KG/HA) INDICADOS NO TRABALHO DE GASQUES ET. AL. 2004 E DA FERRAMENTA - DADOS DO BRASIL EM 2003	142
Tabela 11 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 e DA FERRAMENTA - ÁREA COLHIDA (MIL HECTARES) DE SOJA EM MATO GROSSO	229
Tabela 12 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - PRODUÇÃO (MIL TONELADAS) DE SOJA EM MATO GROSSO	230
Tabela 13 – COMPARAÇÃO DE RESULTADOS E DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - PRODUTIVIDADE (KG/HA) DE SOJA EM MATO GROSSO	231
Tabela 14 – COMPARAÇÃO DE RESULTADOS E DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - ÁREA COLHIDA (MIL HECTARES) DE SOJA NO BRASIL	232

Tabela 15 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - PRODUÇÃO (MIL TONELADAS) DE SOJA NO BRASIL	233
Tabela 16 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - PRODUTIVIDADE (KG/HECTARE) DE SOJA NO BRASIL	234
Tabela 17 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - PRODUÇÃO (MIL TONELADAS) DE ALGODÃO EM CAROÇO NO ESTADO DE MATO GROSSO	235
Tabela 18 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - PRODUTIVIDADE (KG/HECTARE) DE ALGODÃO EM CAROÇO NO ESTADO DE MATO GROSSO	236
Tabela 19 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - PRODUÇÃO (MIL TONELADAS) DE ALGODÃO EM CAROÇO NO BRASIL	237
Tabela 20 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - PRODUTIVIDADE (KG/HECTARE) DE ALGODÃO EM CAROÇO NO BRASIL	237
Tabela 21 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE COSTA, 2016 E DA FERRAMENTA - ÁREA PLANTADA (MIL HECTARES) DE SOJA POR REGIÃO DO BRASIL NO ANO DE 2015	238
Tabela 22 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE COSTA, 2016 E DA FERRAMENTA - PRODUÇÃO (MIL TONELADAS) DE SOJA POR REGIÃO DO BRASIL NO ANO DE 2015	238

LISTA DE QUADROS

QUADRO 1 – INSUMOS DA FASE ANTES DA PORTEIRA	39
QUADRO 2 – DEFINIÇÕES DE DADO, INFORMAÇÃO E CONHECIMENTO . .	45
QUADRO 3 – SISTEMAS DE MEDIÇÃO DE DESEMPENHO	54
QUADRO 4 – INDICADORES FINANCEIROS DO AGRONEGÓCIO	56
QUADRO 5 – INDICADORES NÃO-FINANCEIROS DO AGRONEGÓCIO . . .	57
QUADRO 6 – ELEMENTOS DO DUBLIN CORE	68
QUADRO 7 – POSSÍVEIS FONTES DE DADOS DO AGRONEGÓCIO BRASI- LEIRO	70
QUADRO 8 – TIPOS DE INTERAÇÃO DO USUÁRIO NA VISUALIZAÇÃO DA INFORMAÇÃO	74
QUADRO 9 – ENDEREÇO ELETRÔNICO DE ORGANIZAÇÕES LIGADAS AO AGRONEGÓCIO	85
QUADRO 10 – COLUNAS DA PLANILHA DE DADOS DO AGRONEGÓCIO . . .	87
QUADRO 11 – METADADOS DA FERRAMENTA	90
QUADRO 12 – TRABALHOS UTILIZADOS NA VALIDAÇÃO DOS DADOS FER- RAMENTA	95
QUADRO 13 – DADOS DA AGRICULTURA DISPONIBILIZADOS PELA CONAB	164
QUADRO 14 – DADOS DE AGRICULTURA DISPONIBILIZADOS PELO IBGE .	165
QUADRO 15 – DADOS DE PECUÁRIA DISPONIBILIZADOS PELO IBGE - ABATE	165
QUADRO 16 – DADOS DA AGRICULTURA DISPONIBILIZADOS PELA FAO . .	166
QUADRO 17 – DADOS DE PECUÁRIA DISPONIBILIZADOS PELA FAO - QUAN- TIDADE DE ANIMAIS VIVOS	167
QUADRO 18 – DADOS DE PECUÁRIA DISPONIBILIZADOS PELA FAO - PRO- DUTOS PRIMÁRIOS DE ORIGEM ANIMAL	167

LISTA DE ABREVIATURAS E SIGLAS

BDTD	Biblioteca Digital Brasileira de Teses e Dissertações
BSC	Balanced Scorecard
CSV	Comma-Separated Values
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CNA	Confederação da Agricultura e Pecuária do Brasil
CNPQ	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CONAB	Companhia Nacional de Abastecimento
DB	Data Base
DCMI	Dublin Core Metadata Initiative
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
XML	Extensible Markup Language
FPNQ	Fundação Prêmio Nacional da Qualidade
GI	Gestão da Informação
IBGE	Instituto Brasileiro de Geografia e Estatística
MAPA	Ministério da Agricultura, Pecuária e do Abastecimento
PDF	Portable Document Format
PIB	Produto Interno Bruto
PPGGI	Programa de Pós-Graduação em Gestão da Informação
SciELO	Scientific Eletronic Library Online
SI	Sistema de Informação
SMD	Sistemas de Medição de Desempenho
SIG	Sistema de Informação Gerencial
SMART	Strategic Measurement and Reporting Technique
SQL	Structured Query Language

TB	Tableau de Bord
TI	Tecnologia da Informação
UFPR	Universidade Federal do Paraná
URL	Uniform Resource Locator
VI	Visualização da Informação

SUMÁRIO

1	INTRODUÇÃO	19
1.1	PROBLEMA DE PESQUISA	20
1.2	OBJETIVOS	22
1.2.1	Objetivo Geral	22
1.2.2	Objetivos Específicos	22
1.3	JUSTIFICATIVA	22
1.4	DELIMITAÇÃO DA PESQUISA	27
1.5	ESTRUTURA DO TRABALHO	27
2	REFERENCIAL TEÓRICO	30
2.1	AGRONEGÓCIO	31
2.1.1	Agronegócio no Brasil	34
2.1.2	Cadeia Produtiva do Agronegócio no Brasil	37
2.2	GESTÃO DA INFORMAÇÃO	41
2.2.1	Dado, informação conhecimento	45
2.2.2	Tecnologia da Informação	47
2.2.3	Gestão da Informação e Agronegócio	48
2.3	INDICADORES	50
2.3.1	Sistemas de medição de desempenho	53
2.3.2	Indicadores de Desempenho e Agronegócio	55
2.4	FONTES DE DADOS E INFORMAÇÕES	58
2.4.1	Extração e tratamento de dados	60
2.4.2	Estruturas de Dados	63
2.4.3	ETL e Linguagens de Programação	65
2.4.4	Metadados	67
2.4.5	Fontes de dados do agronegócio	69
2.5	VISUALIZAÇÃO DA INFORMAÇÃO	70
2.5.1	Representações Gráficas	74
3	MATERIAIS E MÉTODOS	81
3.1	MATERIAIS	81
3.2	FERRAMENTAS	82
3.3	MÉTODOS	83
3.3.1	Seleção das fontes de dados	84
3.3.2	Coleta de Dados	85
3.3.3	Tratamento dos dados	86

3.3.4	Definições da ferramenta	88
3.3.5	Hospedagem da ferramenta	91
3.3.6	Comparação de resultados da ferramenta e da literatura	95
4	RESULTADOS	96
4.1	SELEÇÃO DAS FONTES DE DADOS	96
4.1.1	Companhia Nacional do Abastecimento	96
4.1.2	Confederação da Agricultura e Pecuária do Brasil	97
4.1.3	Empresa Brasileira de Pesquisa Agropecuária	97
4.1.4	Instituto Brasileiro de Geografia e Estatística	100
4.1.5	Ministério da Agricultura, Pecuária e Abastecimento	101
4.1.6	Ministério da Indústria, Comércio Exterior e Serviços	101
4.1.7	Organização das Nações Unidas para Agricultura e Alimentação . . .	101
4.2	COLETA DOS DADOS	102
4.2.1	Coleta de dados CONAB	103
4.2.2	Coleta de dados EMBRAPA	104
4.2.3	Coleta de dados IBGE	106
4.2.4	Coleta de dados FAO	109
4.3	TRATAMENTO DOS DADOS	111
4.3.1	Tratamento dos dados CONAB	111
4.3.2	Tratamento dos dados EMBRAPA	114
4.3.3	Tratamento dos dados IBGE	117
4.3.4	Tratamento dos dados FAO	120
4.3.5	Tratamento geral dos dados	123
4.4	DEFINIÇÕES DA FERRAMENTA	124
4.4.1	Gráficos de Linha	125
4.4.2	Métodos Estatísticos e Boxplot	126
4.4.3	Filtros	127
4.4.4	<i>Dashboard</i>	129
4.4.5	Rotina de captação, tratamento dos dados e publicação da ferramenta	130
4.5	HOSPEDAGEM	130
4.6	VALIDAÇÃO DOS DADOS DA FERRAMENTA	131
5	CONCLUSÃO	143
5.1	OBJETIVOS DE PESQUISA	143
5.2	CONSIDERAÇÃO FINAL	145
5.3	SUGESTÕES PARA TRABALHOS FUTUROS	146

REFERÊNCIAS BIBLIOGRÁFICAS	148
---	------------

APÊNDICE A – DADOS DE PRODUTOS DISPONIBILIZADOS PELA CONAB	164
APÊNDICE B – DADOS DE PRODUTOS DISPONIBILIZADOS PELO IBGE	165
APÊNDICE C – DADOS DE PRODUTOS DISPONIBILIZADOS PELA FAO	166
APÊNDICE D – <i>SCRIPT</i> DE COLETA E TRANSFORMAÇÃO DOS DADOS	168
APÊNDICE E – <i>SCRIPT</i> DA FERRAMENTA EM R DE ANÁLISE DOS DADOS	213
APÊNDICE F – COMPARAÇÕES DOS RESULTADOS DE APRE- SENTADOS NO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - SOJA	229
APÊNDICE G – COMPARAÇÕES DOS RESULTADOS DE APRE- SENTADOS NO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - AL- GODÃO EM CAROÇO	235
APÊNDICE H – COMPARAÇÕES DOS RESULTADOS DE APRE- SENTADOS NO COSTA, 2016 E DA FERRAMENTA - SOJA	238

1 INTRODUÇÃO

A informação e sua gestão se fazem essenciais no mundo competitivo de hoje. O agronegócio, um dos principais expoentes da economia do Brasil, responsável pelo setor agrícola e pecuário, também deve fazer uso da transformação de dados em informações, bem como sua gestão posterior.

A agricultura familiar, caracterizada por ser de pequena escala e policultural, transformou-se e modernizou-se após a II Guerra Mundial e fez com que surgisse o termo *agribusiness* de Davis e Goldberg, que cunharam o termo em 1957 (MENDONÇA, 2015). A Teoria da Modernização trouxe uma mudança no panorama da agricultura. Deste modo, a agricultura familiar começa a perder espaço, dando lugar a produções de larga escala em monoculturas, apoiadas na tecnologia de novos maquinários e insumos (BRUM, 1985); (KARNOPP; OLVEIRA, 2012).

Essa modernização fez com que a produção agrícola mundial obtivesse um salto de produção. Essa guinada nos resultados possibilitou ao agronegócio fazer parte do grupo de ramos protagonistas da economia brasileira, contribuindo para o crescimento do Produto Interno Bruto (PIB) (FIGUEIREDO; SANTOS; LIMA, 2012); (ABBADE, 2014). No Brasil, o agronegócio foi responsável por 23% a 24% do PIB no trimestre de julho-agosto-setembro de 2017 (BRASIL, 2017).

Importante ramo de mercado para o Brasil, o agronegócio deve prevalecer-se de uma gestão eficiente e eficaz nas esferas pública e privada para alavancar ainda mais seu crescimento e rendimento. Para isso, o processo de tomada de decisão, caracterizada por Callado e Soares (2014) como percepção de realidade instável com intuito de buscar a melhor solução objetivando estratégias de maneira rápida e eficiente, mostra-se vital para o sucesso.

Na busca da melhor estratégia para os negócios, dados disponibilizados no tempo certo podem ser fundamentais para o sucesso, e esta disponibilização pode auxiliar no cumprimento de obrigações organizacionais e na descoberta de possíveis oportunidades (MOSLEY, 2009)

Devido à alta disponibilidade de dados oriundos do agronegócio, o uso de sistemas de informação se mostra valioso no auxílio da interpretação do grande volume de dados gerados na cadeia produtiva agroindustrial, com objetivo de auxiliar a tomada de decisão (FERREIRA; CAMARGO, 2013).

No agronegócio, a tomada de decisão deve ser, assim como nos demais ramos competitivos de mercado, apoiada em informações confiáveis, advindas de ferramentas que trabalhem com dados relacionados ao negócio, neste caso a produção mostra

indicadores de desempenho valiosos na tomada de decisão (WILDAUER, 2015); (OIKAWA, 2017).

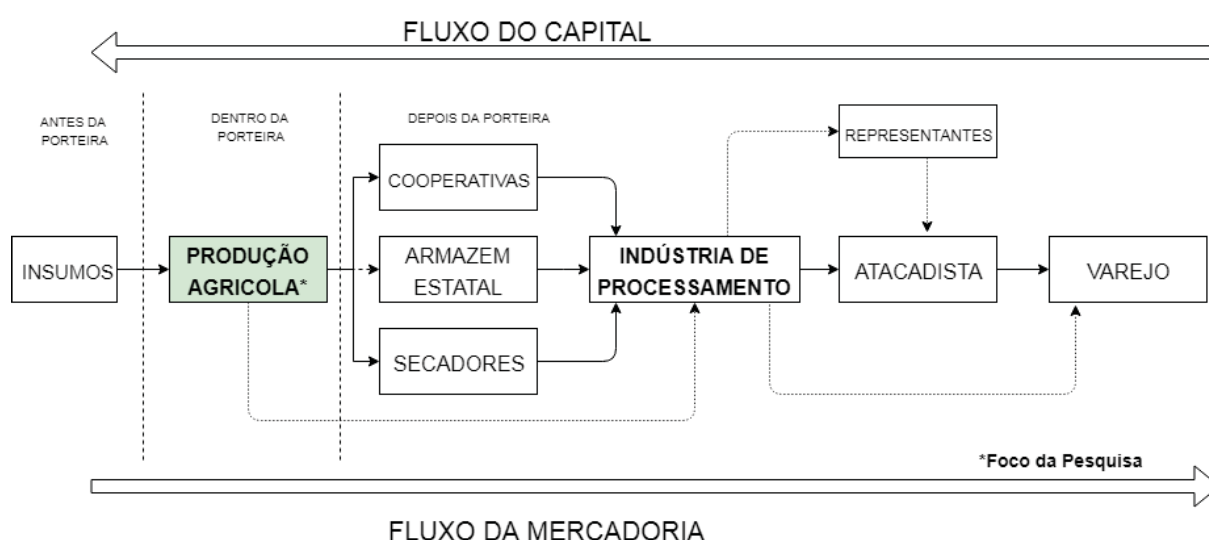
No contexto das empresas, os indicadores se comportam como uma métrica de resultados, com intuito de que esses resultados possam ser comparados com metas estabelecidas anteriormente (CALDEIRA, 2016). A implantação de indicadores de desempenho, pode apontar problemas ou oportunidades dentro de um contexto, oferecendo maior controle da empresa, comunicação de objetivos, motivação dos funcionários e direcionamento de melhoras (FRANCISCHINI; FRANCISCHINI, 2017).

Os indicadores de desempenho cumprem diversas funções, dentre elas, destacam-se, dentro do contexto do agronegócio, a possibilidade de visualização do desempenho de um determinado produtor ou empresa (FERREIRA et al., 2008), e insumo para a formulação de uma estratégia de decisão tomada a partir das informações extraídas dos indicadores (MIRANDA; REIS, 2006).

1.1 PROBLEMA DE PESQUISA

O agronegócio no Brasil é, segundo Welch e Fernandes (2008), um sistema formado pela agricultura, indústria, mercado e finanças. Esses sistemas formam a complexa cadeia produtiva agroindustrial. A Figura 1 demonstra como opera, genericamente, a cadeia do agronegócio, e indica qual o foco da pesquisa.

Figura 1 – OPERAÇÃO GENÉRICA DA CADEIA PRODUTIVA AGROINDUSTRIAL



Fonte: o Autor, adaptado de Zylbersztajn (2000) e Araújo (2013)

Dentro desta cadeia produtiva, autores como Batalha (2001), Zylbersztajn (2000) e Megido e Xavier (2003) citam a existência de três fases, são elas:

a) Antes da porteira: serviços, pesquisa, desenvolvimento e fornecimento de matéria-prima e maquinários necessários se fazem presentes nesta fase.

b) Dentro da porteira: local onde se encontra o fazendeiro ou empresário rural. Abrange atividades produtivas de produtos futuramente vendidos.

c) Depois da porteira: fazem parte desta fase as agroindústrias, comerciantes e distribuidores. Atividades relacionadas com o beneficiamento, distribuição e comercialização dos produtos também fazem parte desta etapa da cadeia.

Dentro do foco da pesquisa, que compreende dados de produção agrícola, ou seja, dentro da porteira, indicadores de produtividade ou de desempenho se fazem essenciais para uma gestão qualificada (OIKAWA, 2017).

Como exemplo de indicadores produzidos na fase dentro da porteira, destacam-se a produção (em toneladas), área plantada e colhida (hectares) e o rendimento por hectare (CONTINI et al., 2006); (MAIA et al., 2015); (GASQUES et al., 2004); (SOUSA; BONJOUR; FIGUEIREDO, 2006).

Os indicadores de desempenho se mostram de grande valor para o mundo dos negócios pelo fato de poderem apontar tendências futuras, o que é de suma importância para as organizações (MIRANDA; REIS, 2006).

Apesar da importância de indicadores, as plataformas governamentais, com exceção da Companhia Nacional do Abastecimento (CONAB), disponibilizam apenas dados brutos de produção agrícola, sem um tratamento prévio que possa auxiliar a tomada de decisão. Segundo Boteon (2004), além da falta de tratamento, estes dados sofrem uma variação em seu conteúdo, precisão e qualidade, devido ao fato de serem oriundas de fontes diferentes.

Devido a complexidade da cadeia produtiva do agronegócio, há muitas vezes dados incompletos e dispersos, caracterizando uma dificuldade de usá-los para fins estratégicos (CORREA, 2010).

Além da falta de indicadores de desempenho na fase “dentro da porteira”, o agronegócio brasileiro sofre com a falta de ferramentas computacionais que forneçam um auxílio na tomada de decisão nesta fase da cadeia a partir de indicadores (CRUVINEL; ASSAD, 2011); (CORREA et al., 2009).

A dificuldade de obtenção de insumos para fim estratégico vem da falta estrutura das informações, ou até da sobreposição de diferentes visões dependendo do interesse na informação (OIKAWA, 2017).

Tendo em vista este panorama, o presente trabalho visa responder a seguinte questão: **como disponibilizar uma ferramenta computacional que colete dados de forma automática a partir de fontes da internet, disponibilizando ao usuário**

indicadores de desempenho de produção do agronegócio brasileiro?

1.2 OBJETIVOS

Os objetivos deste trabalho dividem-se em objetivo geral e objetivos específicos.

1.2.1 Objetivo Geral

Desenvolver uma ferramenta computacional que disponibilize dados de diferentes fontes de dados da web para filtragem dos indicadores e visualização dos mesmos.

1.2.2 Objetivos Específicos

- a) Identificar fontes que disponibilizam dados relacionados ao agronegócio na web.
- b) Propor um modelo de extração automática dos dados de cada uma das bases identificadas.
- c) Construir um modelo de integração dos dados de diferentes bases.
- d) Realizar o mapeamento das variáveis que formarão os filtros da ferramenta para visualização dos dados, os tipos de gráfico de visualização e métodos estatísticos utilizados.
- e) Hospedar a ferramenta.
- f) Comparar resultados da ferramenta com resultados de trabalhos científicos.

1.3 JUSTIFICATIVA

O agronegócio se mostra um dos mais importantes ramos econômicos do Brasil, sendo responsável a uma fatia do Produto Interno Bruto (PIB) que varia geralmente entre 20% e 30% (FERREIRA; CAMARGO, 2013); (ROCHA JUNIOR; BITTENCOURT; RIBEIRO, 2015).

A alta competitividade dentro do âmbito do agronegócio trás a necessidade de sistemas de informação que forneçam informações vitais para o entendimento e aperfeiçoamento de atividades executadas dentro do setor (CALLADO; CALLADO; ALMEIDA, 2008).

Por ser uma cadeia produtiva complexa, o agronegócio gera uma grande variedade de dados que devem ser agrupados e organizados com intuito de de que sejam analisados e demonstrem aos envolvidos (agricultores, empresários, agroindústrias, órgãos do governo) a dinâmica do agronegócio (CORREA, 2010).

A transformação destes dados heterogêneos do agronegócio em informação útil para a tomada de decisão deve ser utilizada na gestão dos negócios. A utilização de ferramentas e técnicas que permitam a geração da informação e o controle dos negócios se faz indispensável (TECH, 2008).

A informação ao alcance de agricultores, investidores, empresas agrícolas e órgãos públicos possibilita a visualização de tendências, que podem determinar rumos nacionais e internacionais para o agronegócio do Brasil, melhorando a competitividade do país frente as demais nações (CORREA, 2010).

Após a obtenção dos dados e realização, através de um sistema de informação, da transformação em informação, é necessário transmiti-la filtrando-a para que seja prática e objetiva no momento de seu uso (OIKAWA, 2017).

Os sistemas de informações que utilizam indicadores de desempenho dividem-se em dois tipos: os que medem desempenho e os que utilizam esses registros de desempenho posteriormente para torná-los informação útil para a tomada de decisão.

A falta de um sistema de medição de indicadores mostra-se como fator negativo dentro de um ambiente empresarial, deixando os envolvidos no processo desorientados e, por fim, culminando na inviabilidade de se fazer com eficácia o gerenciamento (CALLADO; SOARES, 2014).

Já eficácia de uma gestão é pautada, entre outras ações, no uso de um sistema de acompanhamento apoiado nos indicadores de desempenho, que por sua vez dependem de dados (MARTINS, 2006).

Medições existentes possibilitam projetar cenários futuros e tomar melhores decisões no presente, demonstrando que sistemas de indicadores e planos estratégicos são a base da gestão de uma organização (NUTINI, 2015).

Segundo Costa (2011), a capacidade humana tem limites para lidar com grandes quantidades de informação; para isso, o autor cita a utilização de tecnologias da informação no auxílio da interpretação de grandes massas de dados. O uso da visualização da informação através de diagramas e gráficos pode auxiliar a compreensão da informação, sobretudo de indicadores (FRANCISCHINI; FRANCISCHINI, 2017).

A disponibilização de uma ferramenta que possibilite a combinação de dados oriundos de diferentes fontes, que permitam gerar indicadores de desempenho através da livre escolha de filtros, aliada a conceitos de visualização da informação pode ser de grande valia a empresários do setor agrícola, empresas do ramo e até mesmo a esfera governamental, que pode utilizar-se da informação para orientar a tomada de decisão em políticas públicas da agricultura.

Visualizando a importância de ferramentas deste tipo, a Companhia Nacional do

Abastecimento fornece em seu portal eletrônico o Observatório Agrícola, que permite ao usuário a possibilidade de visualizar dados da própria CONAB de maneira interativa, podendo fazer o uso de alguns filtros para visualização da informação através de gráficos.

Apesar de disponibilizar indicadores de desempenho do agronegócio, a ferramenta de visualização da informação da CONAB deixa algumas lacunas em aberto, como por exemplo: utilizar-se somente de dados internos da companhia, não possibilitar ao usuário cruzar dados de diferentes produtos e a falta de mais informação que possa auxiliar na tomada de decisão, como por exemplo, métodos estatísticos.

A Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) também dispõe, em seu endereço eletrônico, de uma ferramenta de visualização da informação de dados do agronegócio brasileiro. A ferramenta da EMBRAPA traz uma abordagem similar a CONAB, de utilização somente de dados internos, onde não é possível cruzar dados de diferentes produtos e a não utilização de métodos estatísticos.

Este panorama deixa a oportunidade de construção de uma ferramenta que contemple os pontos positivos trazidos pela CONAB e contemple também pontos que a ferramenta do órgão governamental acabou não atendendo.

Pode-se afirmar que o agronegócio é um importante ramo econômico do Brasil, e que assim como outros mercados, seu sucesso está diretamente ligado com a informação e sua gestão. Esta afirmação pode-se comprovar também pela quantidade de trabalhos científicos encontrados nas bases que contemplam os termos "agronegócio" e "gestão da informação". Os termos foram pesquisados em qualquer parte do trabalho, com intuito de buscar trabalhos que de alguma forma contemplassem os dois assuntos. Foram consultados os acervos de trabalhos da Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e na Scientific Electronic Library Online (SciELO), os resultados obtidos estão dispostos na Tabela 1.

Tabela 1 – PESQUISA DOS TERMOS "GESTÃO DA INFORMAÇÃO" E "AGRONEGÓCIO"

Termos Pesquisados	BDTD	CAPES	SciELO
"Gestão da Informação" e "Agronegócio"	1209	249	5
Total		1463	

Fonte: o Autor (2018)

Sob o aspecto acadêmico, outros termos relacionados ao trabalho foram pesquisados com o objetivo de demonstrar a relevância dos mesmos em registros acadêmicos. Os trabalhos acadêmicos contemplam a forte relação do agronegócio com os indicado-

res (Tabela 2), que utilizou-se da mesma estratégia de busca da Tabela 1 e das tabelas seguintes apresentadas neste tópico.

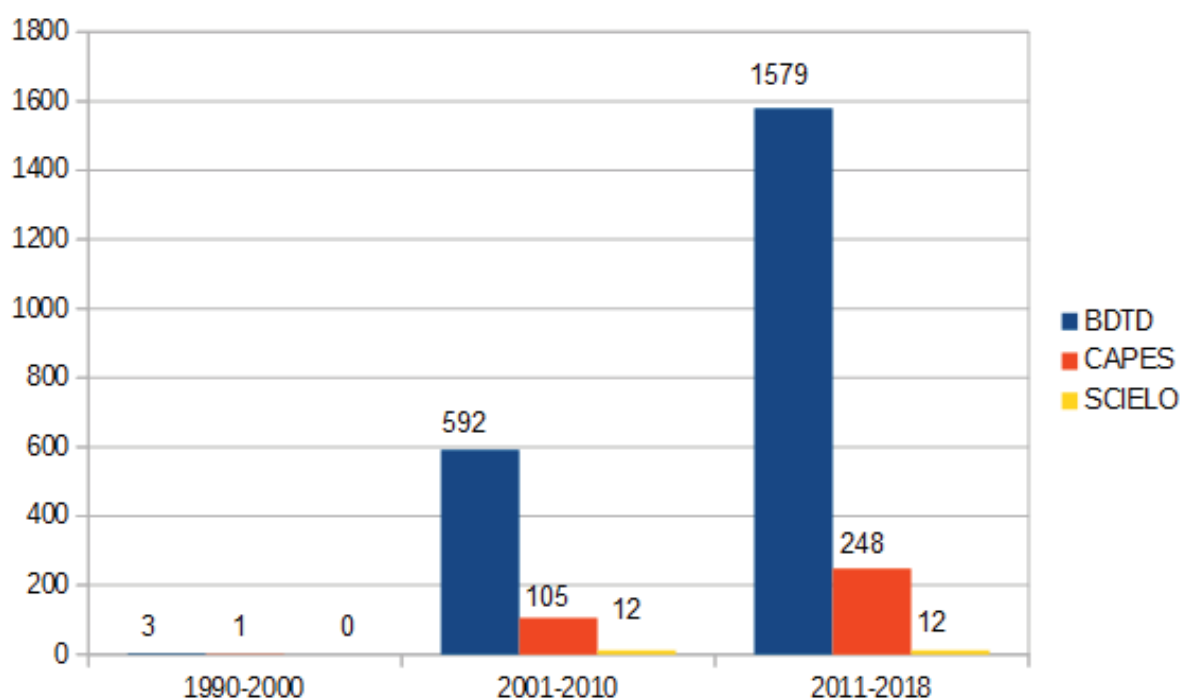
Tabela 2 – PESQUISA DOS TERMOS "AGRONEGÓCIO" E "INDICADORES"

Termos Pesquisados	BDTD	CAPES	SciELO
"Agronegócio" e "Indicadores"	2174	354	24
Total	2552		

Fonte: o Autor (2018)

Sobre indicadores no agronegócio, ressalta-se o aumento de publicações relativas aos assuntos dos anos 90 até o momento (Figura 2).

Figura 2 – DISTRIBUIÇÃO DE TRABALHOS QUE ABORDAM AGRONEGÓCIO E INDICADORES DESDE A DÉCADA DE 90



Fonte: o Autor (2018)

O crescimento do número de trabalhos publicados que contemplam o agronegócio e indicadores passou de quatro trabalho entre 1990 e 2000 para mais de 700 entre 2001 e 2010. Já o período entre 2011 e 2018, que contabilizou cerca de 1800 trabalhos, registrou o aumento de cerca de 250% na produção de trabalhos científicos no contexto se comparado ao período 2001-2010. O aumento no número de produções demonstra que a relação entre indicadores e o agronegócio no Brasil vem ganhando espaço no meio científico

Trabalhos que abrangem questões como a visualização da informação e o agronegócio também foram pesquisados na BDTD, CAPES e SciELO (Tabela 3).

Tabela 3 – PESQUISA DOS TERMOS "AGRONEGÓCIO" E "VISUALIZAÇÃO DA INFORMAÇÃO"

Termos Pesquisados	BDTD	CAPES	SciELO
"Agronegócio" e "Visualização da Informação"	536	29	0
Total		565	

Fonte: o Autor (2018)

Quanto ao foco principal do trabalho da pesquisa, que envolve agronegócio, indicadores e visualização da informação, a pesquisa nas bases retornou os resultados conforme a Tabela 4.

Tabela 4 – RESULTADOS DA PESQUISA SOBRE AGRONEGÓCIO, VISUALIZAÇÃO DA INFORMAÇÃO E INDICADORES

Termos Pesquisados	BDTD	CAPES	SciELO
"Agronegócio" e "Visualização da Informação" e "Indicadores"	535	17	0
Total		552	

Fonte: o Autor (2018)

Com o intuito de refinar ainda mais a pesquisa, buscou-se encontrar os termos "agronegócio", "visualização da informação" e "indicadores" dentro do título, assunto ou resumo, não sendo encontrado nenhuma ocorrência da incidência das palavras juntas em trabalhos. A não ocorrência das palavras juntas em trabalhos nas bases ajuda a reforçar a importância da pesquisa no meio acadêmico.

Quanto à relação com o Programa de Pós-Graduação em Gestão da Informação da UFPR (PPGGI), a pesquisa aborda assuntos contidos na linha de pesquisa "Informação, Tecnologia e Gestão", visando a elaboração de uma ferramenta de auxílio na gestão, sobretudo na tomada de decisão baseada em indicadores de desempenho do agronegócio.

O presente trabalho também se alinha com o projeto de pesquisa do programa, intitulado "Análise de dados e formulação de indicadores de produção do agronegócio do estado do Paraná", que conta com a participação dos professores doutores permanentes do PPGGI Egon Walter Wildauer e José Simão de Paula Pinto.

1.4 DELIMITAÇÃO DA PESQUISA

Conforme (MARCONI; LAKATOS, 2007), cabe-se delimitar uma pesquisa quanto ao seu assunto, extensão e outros fatores.

Quanto ao assunto, ressalta-se que os dados a serem extraídos relacionam-se a fase “dentro da porteira”, mais precisamente dados quantitativos de produção agrícola. Quaisquer outros dados encontrados, e que não fazem parte desta parte da cadeia produtiva agroindustrial, serão ignorados.

Dados financeiros encontrados também não serão considerados. Conforme (BANKER; POTTER; SRINIVASAN, 2000), o uso de indicadores não financeiros mostra-se melhor para previsão de perspectivas futuras.

No que tange a extensão, caso sejam encontradas bases com dados de produção agrícola de outros países juntamente aos do Brasil, os estrangeiros serão descartados, pois não se encaixam no escopo do trabalho, que compreende dados de produção do agronegócio brasileiro.

Com relação a demais fatores, a presente pesquisa limita-se a extrair, forma automática, dados de fontes que as forneçam de maneira pública e gratuita, de diferentes formatos, com diferentes metadados e estruturas em que seja possível a extração automática dos dados via *script*, como por exemplo arquivos .txt, .db, .dbf, .xlsx, .csv, .odf, .xls ou via interface do navegador de internet.

1.5 ESTRUTURA DO TRABALHO

O presente trabalho é dividido em cinco capítulos: introdução, referencial teórico, materiais e métodos, resultados e considerações finais. A pesquisa segue o esquema representado na Figura 3.

Figura 3 – ESTRUTURA DO TRABALHO



Fonte: o Autor (2018)

Introdução - encontra-se neste capítulo a apresentação do tema, o problema de pesquisa os objetivos, os motivos da realização do trabalho que compõem a justificativa, delimitação da pesquisa e, para finalizar, a explanação da estrutura do trabalho. Os tópicos apresentam uma sequência lógica, que visa compor a introdução do trabalho sobre a pesquisa.

Referencial Teórico - esta seção do trabalho aborda assuntos que são pertinentes a presente pesquisa. Agronegócio, gestão da informação, indicadores e visualização da informação são os assuntos que compõem o referencial teórico, dando embasamento ao trabalho através da revisão de literatura dos temas.

Materiais e Métodos - todos os materiais utilizados e a metodologia da cons-

trução do trabalho são indicados neste capítulo. Procedimentos como o de coleta e processamentos dos dados compõem a seção de métodos.

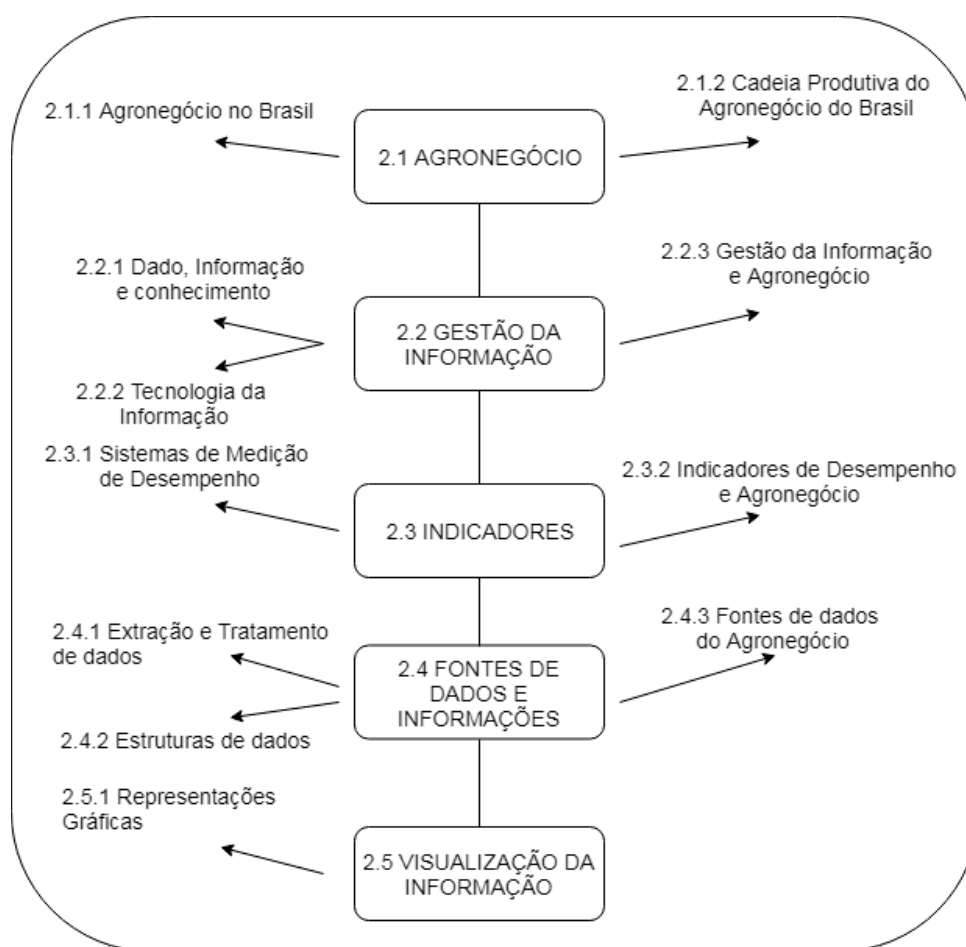
Resultados - indicação dos resultados encontrados na pesquisa. As bases levantadas que fornecem dados do agronegócio, resultados do tratamento dos dados e da disposição dos indicadores via sistema computacional são os focos deste capítulo.

Conclusão - estão contemplados neste tópico a conclusão da pesquisa realizada, bem como a indicação de trabalhos futuros a partir deste.

2 REFERENCIAL TEÓRICO

O referencial teórico (Figura 4) permite embasar o presente trabalho mediante a pesquisa bibliográfica acerca dos temas ligados ao mesmo. O capítulo conta com cinco temas principais.

Figura 4 – ESTRUTURA DO REFERENCIAL TEÓRICO



Fonte: o Autor (2018)

O referencial teórico da presente pesquisa aborda o agronegócio, desde os primórdios da agricultura até seu surgimento. O Tópico ainda contempla a evolução do agronegócio no Brasil e a caracterização da cadeia produtiva agroindustrial brasileira.

Na seção sobre gestão da informação, seus conceitos são abordados, juntamente com a tecnologia da informação e sua relação com o agronegócio. Indicadores trazem consigo suas definições, sistemas de medição de desempenho e por fim indicadores de desempenho dentro do contexto do agronegócio.

Os dois últimos tópicos do referencial teórico abordam as fontes de dados, com estruturas que as compõem e fontes provenientes do agronegócio, e a visualização da informação com sua história, conceitos, vantagens e representações gráficas.

2.1 AGRONEGÓCIO

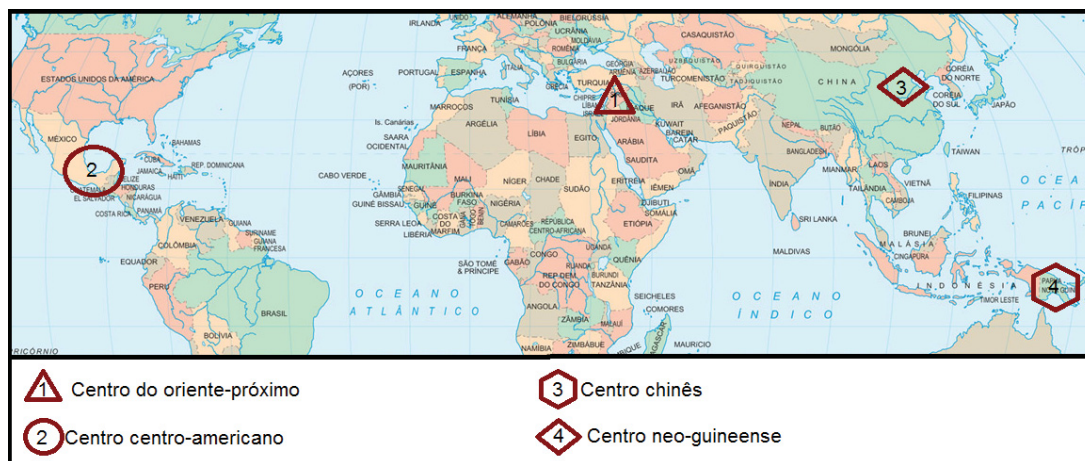
Novos processos de fabricação de instrumentos e polimento da pedra há cerca de 12.000 anos atrás deram início ao último período da Pré-história, chamando de Neolítico, que durou até o surgimento da escrita e da metalurgia (MAZOYER; ROUDART, 2008); o período neolítico, também chamado de idade da pedra polida, é marcado pelo surgimento da civilização, decorrente da mudança do comportamento nômade para o sedentário, domesticação de animais, descobrimento da cerâmica, aprimoramento de instrumentos de corte e o surgimento e desenvolvimento da agricultura (SILVA, 2013); (ALTOÉ; SILVA, 2005); (ABIKO, 2003).

A agricultura teve papel fundamental nas mudanças ocorridas durante o período neolítico, caracterizada como principal recurso de subsistência, ganhou destaque em planícies úmidas, diminuindo o nomadismo característico da época (LIMA, 2016).

O domínio da agricultura fez com que os antigos nômades, agora chamados sedentários, se fixassem perto de rios devido a fertilidade das terras e consumo de água potável. A alta produção agrícola fez com que surgisse o estocamento de alimentos e causou aumento na população, resultando nos primeiros vilarejos e aldeias que posteriormente se tornariam cidades (FABER, 2011).

Segundo Mazoyer e Roudart (2008), quatro centros agrícolas surgidos no período neolítico tiveram destaque (Figura 5).

Figura 5 – GRANDES CENTROS DE AGRICULTURA NO PERÍODO NEOLÍTICO



Fonte: o Autor (2018), adaptado de Mazoyer e Roudart (2008)

- Centro do oriente-próximo - surgiu aproximadamente há 10.000 anos, e foi o mais importante de todos. Situado na Síria-Palestina, contava com a produção de trigo, cevada, lentilha, ervilha e outras leguminosas. A caça era de javalis, coelhos, lebres, etc.
- Centro centro-americano - situado no sul do México, o centro se constituiu entre 9.000 e 4.000 anos atrás, o centro produzia alimentos como milho, abobrinha, abóbora e feijão.
- Centro chinês - o povo chamado Yang Shao se situava nas margens do Rio Amarelo, no norte da China, por volta de 8.500 anos atrás. O cultivo do centro baseava-se em milho, couve, nabo e rami.
- Centro neo-guineense - datado de aproximadamente 10.000 anos, o centro tinha sua produção baseada principalmente em taro.

O desenvolvimento tecnológico, de épocas posteriores ao período neolítico, possibilitou a melhoria em condições higiênicas, o que acabou por reduzir a mortalidade e aumentar a expansão populacional (LIMA, 2016). Para exemplificar, a revolução agrícola e seus avanços entre os séculos XI ao XIII como cultivos em pousio e tração pesada auxiliaram na triplicação da população europeia no período (MAZOYER; ROUDART, 2008).

A revolução agrícola surgiu a partir do século XVIII na Europa Ocidental, e se deu principalmente pela aproximação entre atividades agrícolas e pecuárias, através do aumento da plantação de plantas forrageiras. A rotação de culturas assumiu o lugar da técnica de pousio utilizada anteriormente, a produção aumentou drasticamente, o que acabou abrindo campo para a revolução industrial (EHLERS, 2009).

O período que compreende o final do século XIX e o começo do XX trouxe inovações utilizadas na agricultura como ceifadeiras, semeadeiras e arados. A superprodução obtida através do uso das inovações acabou ocasionando a primeira crise agrícola de superprodução nos anos de 1890 (MAZOYER; ROUDART, 2008).

Segundo Mendonça (2015), a chegada do século XX, trouxe importantes mudanças relacionadas a tecnologia, biologia e química que começam a ser introduzidas na agricultura. Entre as mudanças, destacam-se:

- Início dos 1900 - apesar do primeiro trator ter sido fabricado em 1882, sua aceitação foi lenta e somente em 1917, através do trator criado por Henry Ford, é que os tratores se espalham pelos campos. É o início da mecanização da agricultura (EHLERS, 2009).

- Década de 30 - transformações tecnológicas acabam causando mudanças no padrão de produções, sobretudo em questões ligadas a biologia e química das plantas (MENDONÇA, 2015); (EHLERS, 2009). Outra importante descoberta citada por Ehlers (2009) ocorreu em 1939, com a descoberta das propriedades inseticidas contidas no composto DDT e sua posterior utilização no controle de pragas.
- Década de 50 - após o fim da II Guerra Mundial, há maior intensificação do uso de adubos químicos e mineirais na agricultura (MENDONÇA, 2015); (MAZOYER; ROUDART, 2008).

A aplicação de técnicas aliadas as tecnologias que surgiam no século XX fizeram com que a agricultura manual se tornasse muito menos produtiva frente a motorizada. Milhões de pequenas e médias propriedades acabaram por sucumbir perante grandes e tecnológicas fazendas, que conseguiam até quintuplicar a produtividade (MAZOYER; ROUDART, 2008).

Assim, a relação de produtividade do trabalho entre a agricultura manual menos produtiva e a agricultura motorizada e mecanizada mais produtiva quintuplicou, passando de 1 a 10 no princípio do século a 1 a 500 atualmente (MAZOYER; ROUDART, 2008).

O fim da Segunda Guerra Mundial trouxe a tona a expansão do comércio agrícola mundial. Concomitantemente, a produção de sementes transgênicas e métodos de inseminação artificial são desenvolvidos pela indústria, caracterizando a construção de monopólios industriais na agricultura (MENDONÇA, 2015).

O trabalho de 1957, intitulado *A Concept of Agribusiness* de John H. Davis e Ray A. Goldberg cria o conceito *agribusiness*, difundido no Brasil como agronegócio (MENDONÇA, 2015); (SOUZA; RASIA, 2011); (ARAÚJO, 2013); o termo criado significava a mudança da agricultura, que deixou de ser familiar de subsistência e passou a ser tratada como um negócio. Algumas atividades, que anteriormente eram executadas pelo agricultor, passaram a ser feitas por agentes externos as fazendas (DAVIS; GOLDBERG, 1957).

Dentre estas atividades, Davis e Goldberg (1957) citam a produção de suprimentos agrícolas, como tratores e seu combustível, fertilizantes, inseticidas, estocamento e distribuição do que é produzido dentro das fazendas.

Em suma, o *agribusiness* significa a soma de três setores envolvidos com a agricultura: o de confecção de suprimentos, o da produção dentro da fazenda e o de armazenamento, processamento e distribuição do que é produzido (DAVIS; GOLDBERG, 1957).

Com o passar dos anos, outras características foram sendo acrescentadas ao termo *agribusiness* por autores. Elias (2013) caracteriza o agronegócio como a produção, principalmente, de *commodities*, sendo regulada pela economia de mercado de demandas urbanas e industriais.

Outra forte característica citada por Elias (2013) e Costa (2016) é que o *agribusiness* tem seu foco de comércio voltado à exportação; Costa (2016) ainda afirma que as fazendas participantes da cadeia tem sua produção baseada em monocultura.

2.1.1 Agronegócio no Brasil

O termo *agribusiness* criado por Davis e Goldberg em 1957 demorou algum tempo para chegar ao Brasil. A década de 1960 é marcada, segundo Elesbao (2007), por uma mudança socioeconômica que trás consigo um novo padrão de desenvolvimento rural.

Nos anos 70, o governo federal elegeu o setor agrícola para investimentos. O pós-década de 70 é marcado por condições econômicas, tecnológicas, políticas e culturais que favoreceram a reestruturação do setor produtivo agropecuário brasileiro (SANTOS, 2016).

A extensão territorial, crédito fornecidos e o apoio a pesquisa, sobretudo da Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), são vistos como essenciais para a modernização agrícola brasileira (CONTINI; MARTHA JR., 2010).

Apesar das mudanças ocorridas, o termo agronegócio como é conhecido hoje ainda não existia após a década de 70. A difusão do termo *agribusiness* ocorre nos anos 80 no Brasil, enquanto o vocábulo agronegócio é adotado somente na segunda metade dos anos 1990 (ARAÚJO, 2013).

Assim como nos Estados Unidos, o agronegócio significa no Brasil toda a cadeia que envolve o antes, durante e o depois da produção agrícola. Santos (2016) cita que o agronegócio compreende aspectos administrativos, financeiros, contábeis e logísticos, além de recursos humanos, controle de qualidade e processos, marketing, entre outros.

O agronegócio brasileiro possui características, e, a partir destas, impactos. Como características, o monopólio de terras e políticas agrícolas voltadas para o mercado externo (MENDONÇA, 2015).

Diversos fatores tornam o Brasil destaque na exportação, Ferreira e Camargo (2013) citam a diversidade climática, de solos, e características do país como pontos positivos ao país para o agronegócio. Paiva Junior (2011) também cita a questão dos solos, além dos investimentos em tecnologia e pesquisa, extensão territorial e a integração entre atores da cadeia produtiva do agronegócio.

Outro fator, segundo Figueiredo, Santos e Lima (2012), é que países menos desenvolvidos são estimulados a basear grande parte da economia na produção agrícola. Isso acontece de acordo com os autores pelo fato de atividades agrícolas necessitarem, em média, menor qualificação do que atividades industriais.

Contini e Martha Jr. (2010) indicam quatro motivos pelos quais o Brasil é uma superpotência agrícola: ousadia dos empreendedores, disponibilidade de infraestrutura e terras baratas, pesquisa em tecnologias voltadas para ambientes tropicais e políticas voltadas para o agronegócio, sobretudo crédito rural e preço mínimo de produtos.

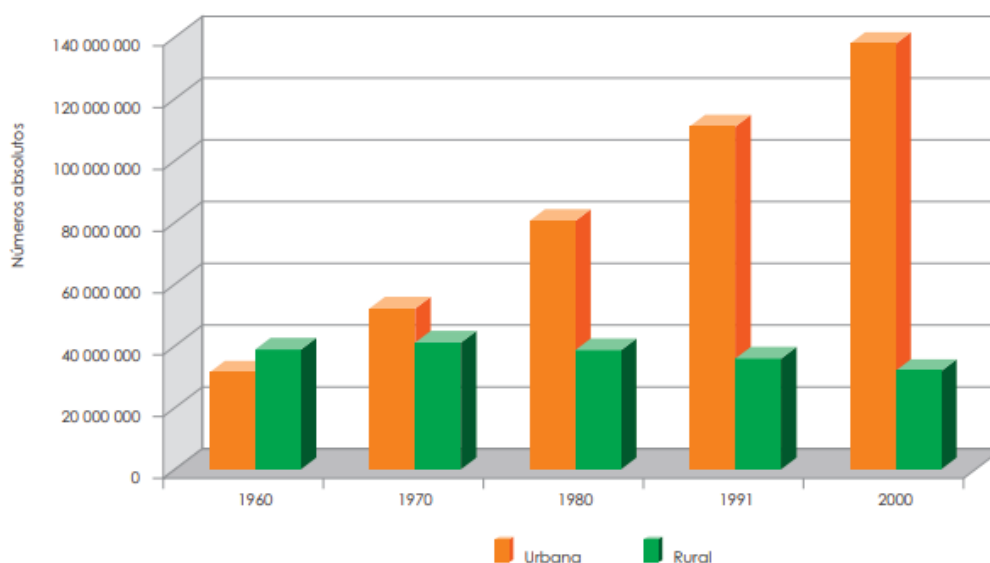
Um dos impactos causados pelas características do agronegócio brasileiro é a reorganização do território do país. Segundo Santos (2016), regiões que recebem grandes investimentos acabam mudando o movimento da cidade, trazendo novos produtos e funções à ela. Mudanças como estas são apontadas no próprio trabalho de Santos (2016), Souza, Santos e Brandão (2015) e Elias (2008).

A diminuição da agricultura familiar de subsistência é outra consequência do agronegócio. Este possível impacto é citado na criação do termo *agribusines* por Davis e Goldberg (1957) e confirmado em trabalhos como de Mendonça (2015), Lima (2016).

Segundo o Instituto Brasileiro de Geografia e Estatística IBGE (2007), 28,3 milhões de pessoas ocupavam o setor rural do país (cerca de 68% da população total) contra 12,9 milhões de pessoas nos centros urbanos (32% do total) na década de 1940.

Com o passar das décadas, estes números foram se alterando, caracterizando êxodo rural, principalmente a partir mecanização do campo nos anos 60 (Figura 6).

Figura 6 – POPULAÇÃO URBANA E RURAL DOS ANOS 60 AOS 2000



Fonte: IBGE (2001)

O último Censo divulgado pelo IBGE (2011), aponta que a população urbana atingiu 84,4% do total (161 milhões de pessoas), enquanto a população rural corresponde a 15,6% do total (29,9 milhões).

A literatura cita diversos outros impactos causados pela expansão do agronegócio no Brasil. Tanto consequências positivas quanto negativas são encontradas.

O principal expoente do agronegócio é sua participação na economia brasileira gerando empregos e renda. Diversos autores como Araújo (2013), Ferreira (2012), Ferreira e Camargo (2013), Figueiredo, Santos e Lima (2012) e Rocha Junior, Bittencourt e Ribeiro (2015) citam a importância do agronegócio no Produto Interno Bruto (PIB) do Brasil, girando em torno de 20% a 30% do total nos últimos 20 anos.

Giordano (2000) cita, como exemplo, o aumento da segurança como ponto positivo causado indiretamente pelo agronegócio, indicando que gestores investem em serviços de vigilância e monitoramento para evitar o roubo de produtos. Essa medida acaba por diminuir a criminalidade nas áreas adjacentes a fazenda monitorada.

Davis e Goldberg (1957) citam, no trabalho que criou o termo *agribusiness*, possíveis desequilíbrios causados pela evolução na relação entre fazendas do agronegócio e camponeses.

O desmatamento e o crescente aumento do uso de agrotóxicos prejudiciais ao meio ambiente também entram na lista de fatores negativos relacionados ao agronegócio (FERREIRA, 2012); (CANUTO, 2004). Como o agronegócio tem seu foco

geralmente ligado a produção de *commodities*, pode também ser considerado uma ameaça a segurança alimentar Costa (2016).

Os chamados *commodities* são produtos caracterizados por Toledo Filho (1991) como:

- Padrão - não sofrem grande variação;
- Pouco perecíveis - geralmente devem ser armazenados e transportados;
- Grande escala de consumo - grande disponibilidade de negociações;
- Fluxo livre e negociado - preço regido pela oferta e demanda;
- Ser matéria-prima ou próximo - não pode depender de um complexo sistema de produção, pois falhas poderiam prejudicar sua disponibilidade

No Brasil, há a prioridade na produção de *commodities*. Segundo o ultimo Censo Agropecuário do IBGE (2006), três *commodities* encabeçam os produtos gerados pelo país: cana-de-açúcar (384 milhões de toneladas), milho (42 milhões de toneladas) e soja (40,7 milhões de toneladas).

Há de se destacar na área pecuária do Brasil, também segundo o Censo Agropecuário do IBGE (2006), a criação de aves (1,2 bilhões de cabeças), bovinos (170 milhões de cabeças) e suínos (32 milhões de cabeças).

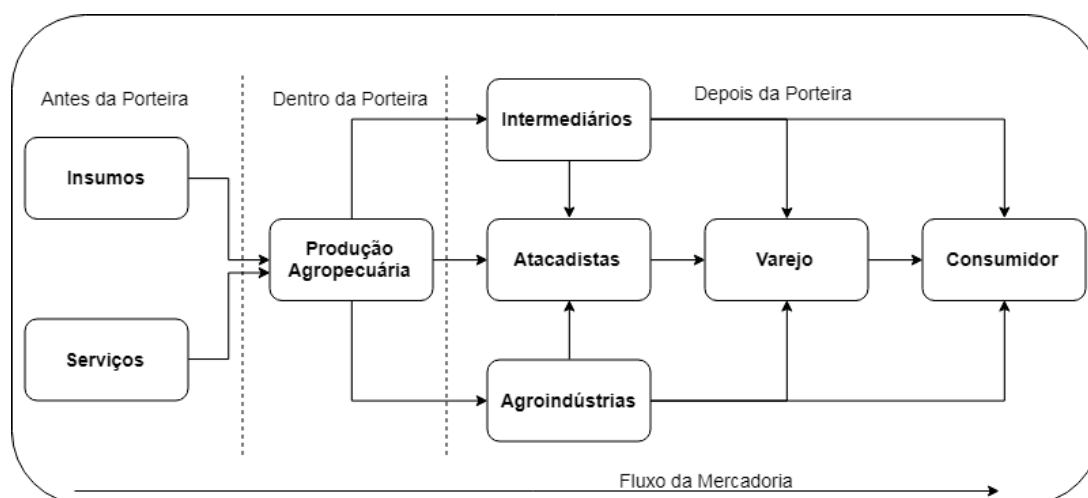
Antes produzidos exclusivamente para alimentação direta da população, o que é produzido atualmente serve muitas vezes de insumo para indústrias de beneficiamento de alimentos, que muitas vezes se situam no próprio meio rural (RIBEMBOIM, 2011). Toda essa produção rural do Brasil depende de um complexo sistema produtivo, que caracteriza o agronegócio como um todo.

2.1.2 Cadeia Produtiva do Agronegócio no Brasil

A cadeia produtiva do agronegócio compreende um complexo sistema, que envolve empresas agropecuárias, indústrias mecânicas, de insumos e sementes, laboratórios de pesquisa, empresas de distribuição, marketing, logística e o comércio em si (ELIAS, 2013).

Todos os envolvidos dividem-se entre três fases da cadeia produtiva agroindustrial (Figura 7): antes da porteira, dentro da porteira e depois da porteira (ZYLBERSZ-TAJN, 2000); (BATALHA, 2001); (MEGIDO; XAVIER, 2003); (CALLADO; CALLADO, 2011b); (SOUZA; RASIA, 2011); (ARAÚJO, 2013).

Figura 7 – CADEIA PRODUTIVA AGROINDUSTRIAL



Fonte: o Autor (2018), adaptado de Araújo (2013)

A primeira fase, denominada "antes da porteira", é composta por dois subsetores. O primeiro deles é responsável por fornecer serviços relacionados ao agronegócio: elaboração e análise de projetos agroindustriais, créditos e financiamentos, capacitação de pessoas, consultoria empresarial especializada, tecnologias da informação, assistências jurídica e à exportação (CALLADO; CALLADO, 2011b); (ARAÚJO, 2013). O segundo subsetor diz respeito aos insumos produzidos para o agronegócio (QUADRO 1).

QUADRO 1 – INSUMOS DA FASE ANTES DA PORTEIRA

Insumo	Descrição	Fonte
Agroquímicos	Agrotóxicos como herbicidas, inseticidas, formicidas, fungicidas e etc.	Araújo (2013), Davis e Goldberg (1957), Ehlers (2009)
Água	Insumo essencial para que a fase "dentro da porteira". Utilizada na irrigação dos solos	Araújo (2013)
Corretivo de solo	Aubos e calcários utilizados na correção de alguma eventual deficiência do solo	Araújo (2013), Callado e Callado (2011), Davis e Goldberg (1957), Mazoyer e Roudart (2008), Mendonça (2015)
Energia	Quaisquer tipo de energia: hidro ou termoelétrica, solar, eólica e etc.	Araújo (2013)
Hormônios	Utilizados para acelerar atividades biológicas de plantas e animais	Araújo (2013), Ehlers (2009), Mendonça (2015)
Maquinário	Tratores, colheitadeiras, semeadeiras, ceifadeiras, caminhões e demais máquinas que dão suporte a produção dentro das fazendas	Araújo (2013), Callado e Callado (2011), Davis e Goldberg (1957) Ehlers (2009)
Materiais genéticos	Os materiais fornecidos as fazendas são: mudas e semente de plantas, e sêmen e óvulo de animais	Araújo (2013), Callado e Callado (2011), Mendonça (2015)
Produtos veterinários	Probióticos, antibióticos, vacinas ecto e endoparasiticidas, estimulantes de apetite e medicamentos	Araújo (2013), Callado e Callado (2011)
Rações	Utilizados na alimentação dos animais; dividem-se em concentrados e volumosos	Araújo (2013), Callado e Callado (2011)

Fonte: o Autor (2018)

Os insumos produzidos na primeira fase da cadeia produtiva do agronegócio, servem como potencializadores à produtividade na fase "dentro da porteira".

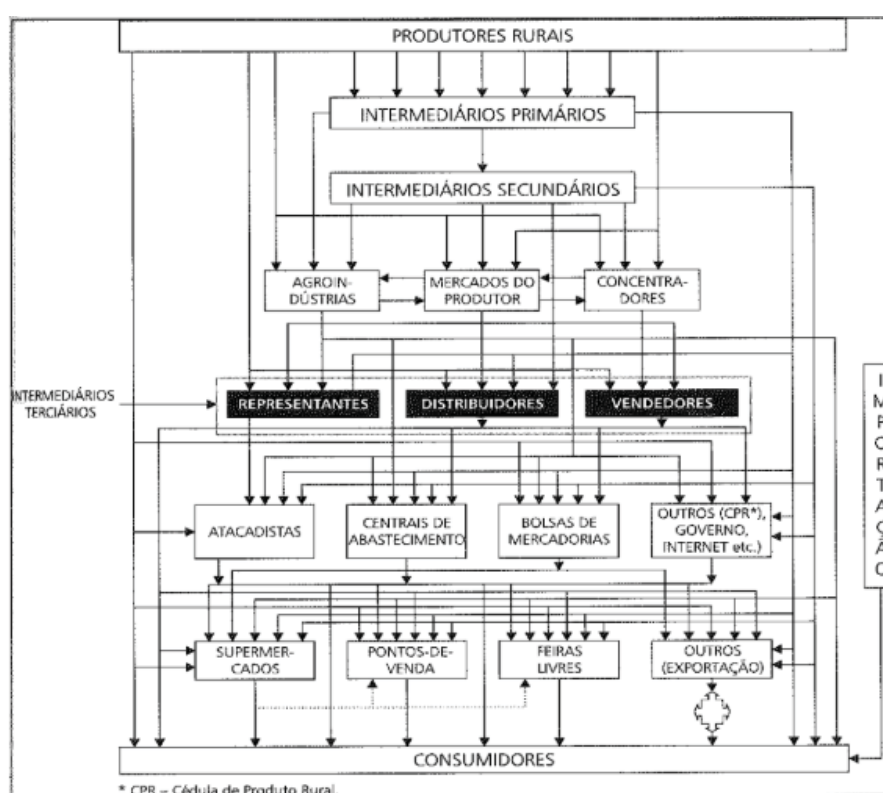
A fase "dentro da porteira" designa as atividades produtivas advindas da exploração econômica de fatores produtivos dentro das fazendas. Entre os fatores, destacam-se as atividades agrícolas e pecuárias (CALLADO; CALLADO, 2011b).

Segundo Araújo (2013), alguns aspectos ligados à produção agrícola são o ciclo

vegetativo (plantas anuais, perenes e semiperenes), preparação dos solos, viveiros e mudas, o plantio, colheita e pós-colheita. Já no segmento pecuário, o autor aponta os aspectos de sistema de criação (intensivos, semi-intensivos e extensivos) e manejo do rebanho.

Após a produção, é necessário que os produtos sejam inseridos no mercado. A última fase, chamada "depois da porteira" abrange essa inserção através da distribuição e comercialização (Figura 8) dos produtos agroindustriais (CALLADO; CALLADO, 2011b).

Figura 8 – CADEIA PRODUTIVA AGROINDUSTRIAL



Fonte: Callado e Callado (2011b)

O complexo processo de comercialização dos produtos no setor "depois da porteira" é dividido em oito níveis por Araújo (2013).

- **Nível 1 - Produtores rurais** - Podem ofertar produtos diretamente para qualquer outro nível, mas geralmente fixam-se em um dos níveis, geralmente intermediários primários;
- **Nível 2 - Intermediários** - Pessoas ou empresas que compra dos produtores rurais e repassam para outros níveis;

- **Nível 3** - Agroindústrias, mercados dos produtores e concentradores - as agroindústrias trabalham com beneficiamento, processamento ou transformação dos produtos. Mercados dos produtores são locais onde diversos agropecuaristas concentram suas produções, recebem informações sobre mercado e demais serviços, são utilizados para diminuir a força de intermediários. Os concentradores são intermediários de grande porte;
- **Nível 4** - Representantes, distribuidores e vendedores - não possuem produto, apenas os repassam para serem comercializados;
- **Nível 5** - Atacadistas, governo, internet, centrais de abastecimento e bolsas de mercadorias - Os atacadistas são grandes empresas que adquirem produtos de todos os níveis abaixo. As centrais de abastecimento funcionam semelhantemente aos mercados produtores, tentam aproximar o produtor rural do consumidor. A bolsa de mercadorias trabalha na oferta de serviços de pregão de produtos agropecuários. O governo, através da Companhia Nacional de Abastecimento (CONAB) trabalha na parte legislativa e econômica do setor. A internet também conta com venda de produtos agropecuários;
- **Nível 6** - Mercados, feiras, pontos de venda e exportação - segmentos de mercado que fazem contato direto com o consumidor e também realizam exportações;
- **Nível 7** - Consumidores - parte final da fase "depois da porteira". É o nível mais importante, pois tudo que é realizado anteriormente tem como objetivo atender ao consumidor;
- **Nível 8** - Importação - a importação de produtos interfere na comercialização interna. Produtos importados geralmente percorrem caminhos parecidos, iniciando seu ciclo geralmente no nível 3.

2.2 GESTÃO DA INFORMAÇÃO

A década de 1950 está ligada ao desenvolvimento dos computadores e a capacidade dos mesmos ao permitir captura e armazenamento de informações em grande escala. Esta capacidade aguçou o interesse de empresas de se planejar e controlar o desempenho organizacional (DAVENPORT; PRUSAK, 1997).

A forma de extração de informações através de tecnologias vem, segundo O'Brien (2004), justamente dos anos de 1950 a 1960, com os primeiros sistemas de processamento eletrônicos de dados; entre os anos 60 e 70, os primeiros registros de apoio à decisão: sistemas de informação gerencial que contavam com relatórios administrativos para o suporte a tomada de decisão.

A década 70, foi marcada por ser um capítulo importante da evolução do capitalismo ao emergiu a necessidade da adoção de novas estratégias pelas empresas para combater o declínio nos lucros. Uma das estratégias utilizadas foi a adoção de inovações tecnológicas, o que acabou por causar o surgimento da chamada revolução da tecnologia da informação (CASTELLS, 2005).

Apesar de não se saber ao certo quando surgiu, podemos dizer a GI como conhecemos hoje foi colocada mais em foco a partir da década de 90, juntamente com o surgimento do conceito de Gestão do Conhecimento, que engloba as gestões da informação e de pessoas (SVEIBY, 2003).

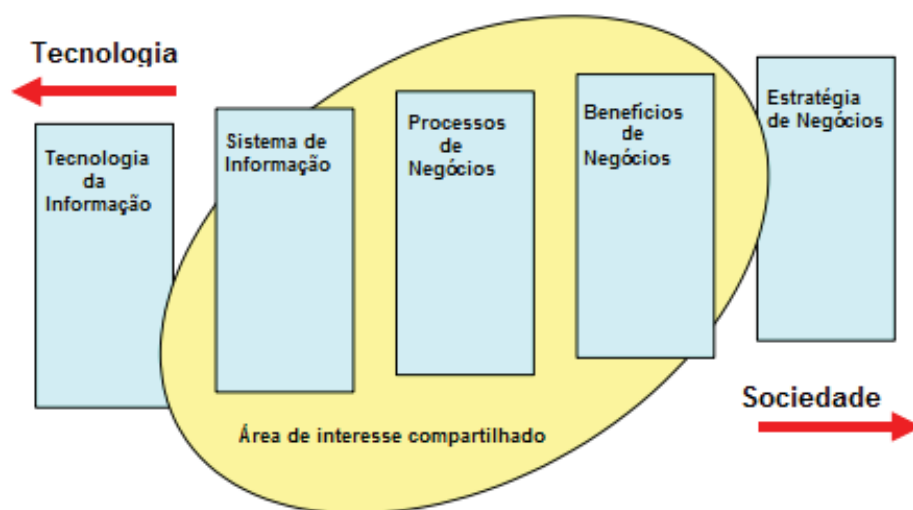
Após o surgimento da Era da Informação, a década de 90 trouxe a popularização dos computadores pessoais e da internet, juntamente com um cenário econômico em que se exigia flexibilidade e inovação, fazendo com que organizações buscassem na informação a vantagem para se manterem competitivas (PINTO, 2013).

A capacidade organizacional de criar, manter e disponibilizar a informação para a pessoa correta é o que define a gestão da informação (GI) (CHOO, 2003). Davenport e Prusak (1997) caracterizam a GI como um conjunto estruturado de atividades correspondentes a captura e distribuição da informação e conhecimento dentro de um ambiente organizacional.

Maravilha-Lopes (2013) afirma que a GI é uma atividade que visa regular a informação, as tecnologias e os utilizadores, tudo isso aplicando técnicas de gestão, afim de processar e disponibilizar informações pertinentes e atualizadas para auxílio na tomada de decisão e atingimento de objetivos.

Cinco áreas de conhecimento compõem, segundo Bytheway (2004, p. 8) a GI: tecnologia da informação, sistemas de informação, processos de negócios, benefícios empresariais e estratégia de negócios (Figura 9).

Figura 9 – ÁREAS DE CONHECIMENTO DA GESTÃO DA INFORMAÇÃO



Fonte: Bytheway (2004)

A tecnologia da informação é a parte de *hardware* e *software* que compõem um sistema de informação, que por sua vez, também é composto por humanos que realizam processos e atividades organizacionais a fim de gerar benefícios e estratégias a serem utilizadas pela empresa (BYTHEWAY, 2004).

A gestão da informação é apoiada sobre um sistema de informação (SI). De-coster e Sun (2014) caracterizam um SI como um conjunto de partes ou elementos inter-relacionados para realizar a coleta, processamento, distribuição e disseminação de informações utilizadas no suporte a decisão, tudo isso de forma organizada e coordenada.

A tomada de decisão dentro de uma empresa é caracterizada por Callado e Moraes Filho (2011, p. 25) como "escolha da opção dentro cursos alternativos que melhor se enquadre em seus interesses"; os autores ainda afirmam que a opção escolhida é influenciada por previsões de resultados futuros geradas pelas empresas.

A informação acerca de pontos positivos e negativos do ambiente organizacional interno e externo juntamente com a habilidade de saber utilizá-la pode ser um diferencial estratégico das empresas (SOARES, 2011).

Empresas que tomam suas decisões não serem pautadas em informações tem desvantagem frente as que se utilizam da informação. Todos os tipos de tomada de decisão dependem de informação, o que os diferencia é a forma como as buscam e as utilizam (LOUSADA; VALENTIN, 2011).

O tipo da tomada de decisão pode ser no classificada em um dos quatro

modelos indicados por Choo (2005, p. 211): racional, guiada por regras e procedimentos predefinidos com foco nos objetivos; político, onde se há conflitos nos objetivos, e a decisão é tomada por influência de envolvidos; processual, onde existem vários tipos de ações possíveis e é escolhida uma delas e o modelo anárquico, que é caracterizado por decisões onde objetivos e processos são ambíguos.

A tomada de decisão passa pelas fases de se selecionar uma ocasião para se tomar uma decisão, inventar, desenvolver e analisar possíveis resultados de ações traçadas e selecionar uma delas (SIMON, 1960).

Cabe a informação seguir regras e rotinas organizacionais. Entre as características da informação para a tomada de decisão, destacam-se a disponibilização de alternativas e preferências para o tomador da decisão, bem como a informação seguir regras, rotinas e satisfazer a estratégia organizacional (CHOO, 2005).

Para melhor suporte na tomada decisão se faz necessária a qualidade da informação. Conforme Chaffey e Wood (2004, p. 23) se não há qualidade na informação, os processos são feitos de maneira ineficiente e a informação gerada não será confiável, podendo direcionar a decisões equivocadas.

Segundo O'Brien (2004, p. 15), a qualidade da informação reside nas perspectivas de tempo, conteúdo e forma:

- Tempo - conforme a perspectiva de tempo a informação deve ter:
 - Prontidão - ser fornecida assim que necessária;
 - Aceitação - estar atualizada;
 - Frequência - ser entregue quantas vezes forem necessárias;
 - Período - podem ser de tempos passados, presentes ou futuros.
- Conteúdo - sob a ótica de conteúdo a informação deve conciliar:
 - Precisão - estar isenta de erros;
 - Relevância - deve estar relacionada a uma situação específica e para um receptor específico;
 - Integridade - tudo que for necessário deve ser disponibilizado;
 - Concisão - mostrar apenas o que é necessário;
 - Amplitude - ter alcance grande ou pequeno, com foco interno ou externo;
 - Desempenho - demonstrar o desempenho de atividade concluídas ou em progresso.
- Forma - a forma da informação deve contemplar:

- Clareza - forma de fácil compreensão;
- Detalhe - ser de forma resumida ou detalhada;
- Ordem - organizada em uma sequência predeterminada;
- Apresentação - forma narrativa, numérica, gráfica, etc;
- Mídia - documentos impressos, monitores, etc.

Os sistemas de informação deve ser estruturados para cumprirem as tarefas de armazenamento, organização e distribuição da informação para a tomada de decisão , tudo isso de maneira integrada dentro da organização (SILVA, 2017).

Os sistemas de informação eficientes, pautados em dados, informação e conhecimento, permitem a propagação da informação no ambiente organizacional e, conseqüentemente, auxiliam a tomada de decisão (ANGELONI, 2003).

2.2.1 Dado, informação conhecimento

Para o suporte à tomada de decisão, a gestão da informação faz uso de seus três principais componentes: dados, informação e conhecimento. A literatura apresenta várias definições de vários autores para os componentes da GI (QUADRO 2).

QUADRO 2 – DEFINIÇÕES DE DADO, INFORMAÇÃO E CONHECIMENTO

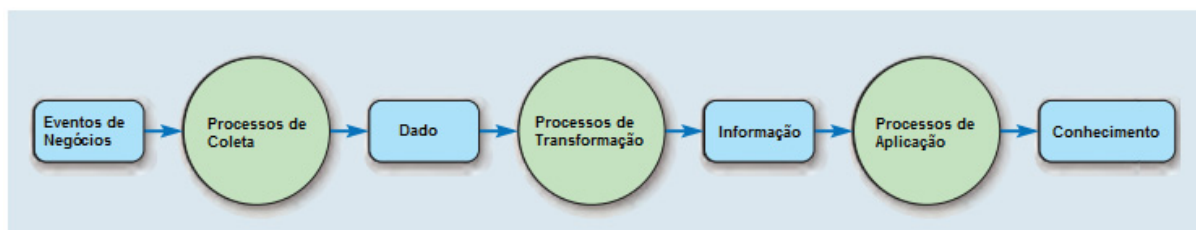
Dado	Informação	Conhecimento
Fatos objetivos e discretos sobre algum evento (CHAFFEY; WOOD, 2004, p. 21)	Dados organizados com significado relevante para tomada de decisão (CHAFFEY; WOOD, 2004, p. 21)	Dados e informações adicionais a opinião de especialistas para se tornar um recurso de tomada de decisão (CHAFFEY; WOOD, 2004, p.21)
Fatos ou observações crus, sem um tratamento (O'BRIEN, 2004)	Dados convertidos para um contexto significativo e usuários específicos (O'BRIEN, 2004)	Domínio cognitivo que pode ser compartilhado (CHECKLAND; HOLWELL, 2006)
Sequência de símbolos quantificados ou quantificáveis (SETZER, 1999)	Dados dotados de relevância e propósito (DRUCKER, 1988)	Informação combinada com experiência, contexto, reflexão e interpretação (DAVENPORT; DE LONG; BEERS, 1998)
Entidade mais primitiva e fundamental da GI (THOW-YIK, 1994, p. 651)	Agregado de dados capaz de transmitir uma mensagem significativa (THOW-YIK, 1994, p. 651)	Combinação seletiva de uma grande massa de informações acumuladas (THOW-YIK, 1994, p. 651)

Fonte: o Autor (2018)

Os dados são transformados em informação através dos sistema de informação contido no âmbito, para que possam ser utilizados posteriormente na tomada de decisão

(CHAFFEY; WOOD, 2004). O procedimento chamado "processamento de dados", onde são agregados manipulados e organizados para que seu conteúdo possa ser avaliado e após isso, serem colocados no contexto adequado para o usuário (O'BRIEN, 2004); fechando assim o modelo de transformação de processo indicado por Chaffey e Wood (2004, p.22)(Figura 10).

Figura 10 – PROCESSO DE TRANSFORMAÇÃO DA INFORMAÇÃO



Fonte: Chaffey e Wood (2004)

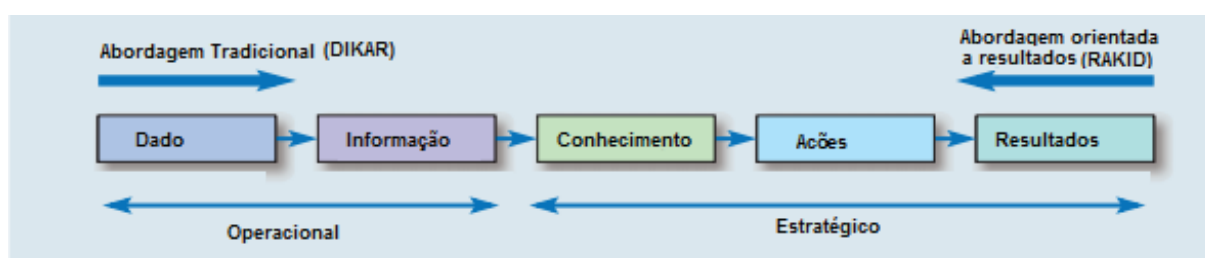
A literatura ainda apresenta modelos que contemplam mais fases além do conhecimento. Os modelos DIKAR e RAKID (Figura 11) são vistos em trabalhos como de Chaffey e Wood (2004), Ward e Peppard (2002).

Além de contemplar a transformação de dados em conhecimento através da informação, o modelo DIKAR (*Data, Information, Knowledge, Action, Results*) leva em conta também as ações tomadas em decorrência do conhecimento obtido e os resultados posteriores as ações (CHAFFEY; WOOD, 2004).

O modelo pode ser utilizado para entender os processos organizacionais e é a abordagem mais utilizada com as tecnologias de informação (WARD; PEPPARD, 2002).

Já o modelo RAKID tem a perspectiva contrária ao DIKAR, onde a abordagem é orientada a resultados. No RAKID é possível selecionar requisitos de dados e informações para fornecerem resultados específicos (CHAFFEY; WOOD, 2004).

Figura 11 – MODELO DIKAR E RAKID

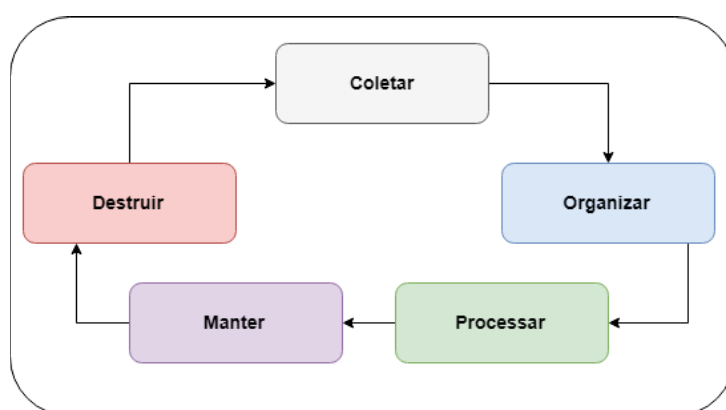


Fonte: Chaffey e Wood (2004)

As ligações entre as fases demonstram as competências da organização, representando atividades onde se tem incremento de valor via procedimentos, sistemas, ações, habilidades, etc (OLIVEIRA et al., 2015).

A gestão da informação competente envolve também o gerenciamento do ciclo de vida da informação (Figura 12), de sua coleta até sua destruição, sendo que alguns passos podem se repetir (CHAFFEY; WOOD, 2004).

Figura 12 – CICLO DE VIDA DA INFORMAÇÃO



Fonte: Chaffey e Wood (2004)

A fase de coleta remete ao registro dos dados, enquanto o ciclo da organização corresponde ao agrupamento dos dados; o processamento compreende a análise; a atualização e revisão da qualidade contemplam a manutenção e a destruição trata da eliminação de informações que já não são mais necessárias (CHAFFEY; WOOD, 2004).

Existem modelos de ciclo de vida da informação um pouco diferente do apresentado por Chaffey e Wood. Marchand, Kettinger e Rollins (2002, p. 1) não indica a fase de destruição da informação, substituindo-a pela detecção de dados, precedendo a etapa de coleta.

O ciclo de vida da informação é dependente do uso de uma tecnologia da informação, juntamente com processos organizacionais internos.

2.2.2 Tecnologia da Informação

A tecnologia da informação, conforme apontado por (BYTHEWAY, 2004), é um dos pilares da gestão da informação, ajudando a compor os sistemas de informação e facilitando a busca de estratégias de negócios empresariais.

Desde a década de 90, o uso da tecnologia é considerado um dos fatores responsáveis pelo sucesso, sobrevivência e competitividade das corporações no mundo

dos negócios (YOUNG, 1992).

O mercado exige aumento da velocidade do dinamismo e execução de processos, e a tecnologia se apresenta a disposição para auxiliar nestas tarefas, mostrando-se um fator de sucesso determinante no mundo dos negócios (WAACK, 2000).

A vantagem competitiva que pode ser fornecida pelas tecnologias fez com que administradores enxerguem a TI como parte do planejamento estratégico da empresa, tratando-a como uma parceira da organização. A TI deve estar alinhada com a estrutura e cultura organizacional, com intuito de diminuir a resistência de colaboradores e aumentar a capacidade estratégica (THEODOROU, 2005).

Como objetivos da tecnologia da informação estão: facilitação da aquisição, processamento, armazenamento e disseminação da informação dentro do âmbito organizacional (WARD; PEPPARD, 2002);

A TI, através de seus componentes de *hardware*, *software* e rede, pode atuar desde o setor mais operacional da empresa até o mais estratégico, dando suporte as operações, processos, inovações e tomada de decisão ((MARCHAND; KETTINGER; ROLLINS, 2002)).

O uso de tecnologias em ambientes organizacionais dentro do contexto de gestão acabam por induzir a produtividade das empresas (CASTELLS, 2005); (DERY; SAMSON, 2005). A utilização de tecnologia da informação pode auxiliar ações estratégicas de uma organização na adição de novos mercados e clientes, controle de riscos, redução de gastos e criação de novos produtos, ideias e serviços (CHAFFEY; WOOD, 2004).

O avanço de tecnologias de inteligência artificial potencializam ainda mais as tecnologias de informação no mercado; estes sistemas geralmente se apoiam em redes neurais e lógica difusa ou *fuzzy* (BEGG, 2005).

Dentro do ramo do agronegócio, pode ser considerada uma vantagem competitiva a escolha da tecnologia, nível de produção e métodos de comercialização feita através de ferramentas que possam auxiliar o gestor na tomada de decisão (SOARES, 2011).

2.2.3 Gestão da Informação e Agronegócio

A chamada Era da Informação, surgida nos anos 90 tem como objetivo dar ênfase em questões como: clientes, competitividade, globalização, produtividade e qualidade (CHIAVENATO, 2004).

O agronegócio, assim como outros ramos, também deve dar ênfase nestas questões, pois segundo Callado e Callado (2011b, p. 1), o agronegócio está inserido

em um meio econômico complexo e diversificado.

Dentro do agronegócio, as empresas rurais são classificadas quanto a seu porte como: pequenas, que utilizam-se de cadernetas para controle financeiro do que se tem a pagar e a receber; médias, onde há registros financeiros de entrada, saída, transações efetuadas, de contas correntes e controle de funcionários; e grandes, que contam com a contabilidade financeira minuciosa e completa (CALLADO; MORAES FILHO, 2011).

O gestor rural tem como objetivos: planejar, controlar, decidir e avaliar os resultados visando lucros, trazer motivação e bem estar de colaboradores (SANTOS; MARION; SEGATTI, 2009). De acordo com Callado e Moraes Filho (2011, p. 26), o gestor ainda deve estar pronto para identificar possíveis ameaças e oportunidades que possam surgir.

No mundo competitivo como o do agronegócio onde sobrevivem apenas os diferentes ou os que estão em melhor posição estratégica, ser um bom estrategista é essencial para o sucesso empresarial (SOARES, 2011).

Duas perspectivas fazem parte da vida do gestor rural: processos produtivos, que estão relacionados a questões internas e as atividades comerciais, ligadas ao ambiente externo. A gestão deve levar em conta elementos financeiros, produtivos e sociais (CALLADO; MORAES FILHO, 2011). Paiva Junior (2011, p. 50) ainda cita que o gestor rural deve ter capacidade de agir de forma rentável com relação aos negócios, sempre com responsabilidade no âmbito socioambiental.

A importância da utilização de sistemas apoiados em informação é apontado por (CALLADO; MORAES FILHO, 2011), que afirma que as informações geradas podem guiar, sobretudo, o marketing dentro da cadeia do agronegócio.

A competitividade dentro do agronegócio é composta por três capacidades, que são de produção/tecnológica, de inovação e de coordenação. A última delas refere-se a capacidade de processar e utilizar informações focadas na estratégia, controle e reação a mudanças no meio ambiente (JANK; NASSAR, 2000).

As práticas gerenciais utilizadas no agronegócio priorizam a parte operacional e mercadológica, com intuito de agregar valor a serviços e produtos, tendo como consequência inserção, consolidação e expansão no mercado (CALLADO; MORAES FILHO, 2011).

A informação necessária para a tomada de decisão num momento de incerteza no agronegócio pode vir da análise estatística sobre seus dados, que é responsável por analisar fenômenos quantitativos através de seus métodos. Além do suporte à decisão em momentos de incerteza, a estatística pode auxiliar na descrição de um conjunto de dados e na explicação e previsão de certos fenômenos (MÖLLER, 2011)

Métodos da estatística descritiva, como cálculo de média, mediana, desvio padrão, variância e valores mínimos e máximos compõem a estatística descritiva, responsável por descrever um conjunto de dados (PINHEIRO et al., 2012).

Tukey (1977) apresenta métodos de análise de estatística descritiva com resultados apresentados em gráficos, sobretudo *boxplots* e gráficos de barras, tendo como objetivo facilitar a análise dos resultados obtidos da base de dados .

A estatística descritiva auxilia no processo de tomada de decisão das esferas pública e privada através da explicação dos fenômenos reais quantitativamente por meio de tabelas e gráficos (MÖLLER, 2011).

2.3 INDICADORES

Dentro da gestão empresarial, a avaliação do desempenho de uma organização se mostra atividade fundamental (NASCIMENTO et al., 2011), e para isso, os indicadores tornaram-se uma ferramenta gerencial e uma medida estratégica de sobrevivência das empresas (COELHO et al., 2008).

O uso de indicadores se propagou nos anos 90, de acordo com Neely (1999, p. 210), algumas razões como a mudança nas demandas, natureza mutável do trabalho, aumento de concorrência e o poder da tecnologia da informação ajudam a justificar a proliferação do uso.

A quantificação de como são realizadas as atividades é o que define indicadores de desempenho, tendo como um dos objetivos compará-los com metas previamente estipuladas (MACARTHUR, 1996).

Os indicadores são medidas diretas ou indiretas da qualidade, que tem o intuito de avaliar a situação atual e apontar tendências futuras de um determinado ambiente. São capazes de demonstrar a degradação de um sistema, bem como alertar sobre potenciais problemas (TOCCHETTO; PEREIRA, 2008).

Para Van Bellen (2004, p. 6), os indicadores tem o intuito de agregar e quantificar informações, objetivando maior aparência na significância, além de simplificar informações sobre processos complexos.

Os indicadores de desempenho tem como objetivo auxiliar a compreensão organizacional, e a partir disso, permitir que as decisões tomadas esteja alinhadas com a estratégia da empresa (PORTO; ESTRADA, 2004).

É imprescindível que os envolvidos nas organizações sejam treinados a interpretar os indicadores, e a partir disso, tomar decisões (MARTINS; COSTA NETO, 1998).

Conforme (CALDEIRA, 2016), para que os indicadores sejam úteis para a gestão devem possuir algumas características como:

- Pertinência para a gestão - o resultado apresentado pelo indicador deve ser útil para o gestor;
- Credibilidade de resultado - os dados inseridos para a geração dos indicadores devem ser confiáveis, caso contrário comprometem a credibilidade dos indicadores;
- Esforço aceitável para se chegar no resultado - o esforço para se adquirir os indicadores deve ser aceitável. O custo para se obter a informação não pode ser maior do que o valor que ela fornece;
- Simplicidade - os indicadores gerados devem ser fáceis de se interpretar;
- Algoritmo de cálculo simples e automático - o procedimento para se chegar ao resultado medido deve ser simples, mais automático possível e evitar intervenção humana;
- Fonte de dados interna e auditável - as bases com os dados devem contar com fácil acesso, preferivelmente dentro da organização e que possam ser auditáveis;
- Frequentes e atualizados - a disponibilização deve ter frequência maior ou igual a necessária pelos gestores, estando sempre atualizados;
- Proteção contra efeitos externos e perversos - além de não poderem ser afetados por efeitos externos que podem esconder a verdade dos resultados, os indicadores não podem trazer efeitos negativos para dentro de uma organização;
- Possibilitar *benchmarking* - a comparação dos desempenhos de diferentes atividades, projetos, setores e empresas é útil para se incentivar a concorrência e promover a melhora contínua;
- Possibilidade de meta - é importante que existam metas, para se diagnosticar a distância dos resultados obtidos com os considerados ideais.

O alinhamento de indicadores tem papel fundamental nas organizações, pois é preciso monitorar e controlar resultados obtidos para que se possa alterar o planejamento atual estabelecido ou definir nova rota estratégica (PORTO; ESTRADA, 2004).

Quanto ao alinhamento, Callado e Soares (2014, p. 273) citam que os indicadores, além de serem compreensíveis e controláveis, devem estar alinhados à estratégia

organizacional, fazendo com que os recursos humanos produzam de maneira coerente com a estratégia traçada.

No que diz respeito a tomada de decisão, os indicadores de desempenho atuam como medida de avaliação, fornecendo *feedback* de atividades e estimulando pessoas em suas tarefas (CALLADO; SOARES, 2014).

Este tipo de indicador tem sido utilizado para gestores avaliarem o desempenho, e a partir disso, auxiliarem a tomada de decisão (CALLADO; CALLADO, 2011a).

Diversas características definem e dividem os indicadores de desempenho. Paula e Ichikawa (2002, p. 1) classifica indicadores de qualidade e de produtividade, onde os de qualidade dizem respeito a satisfação de clientes, os de produtividade tratam da medição do desempenho de processos e utilização de recursos disponíveis.

Sink e Tuttle (1993) separa os indicadores quanto a visibilidade (identificação de pontos fortes e fracos), controle (possibilidade de controlar o desempenho frente a padrões estabelecidos anteriormente), e melhoria (demonstração dos resultados comparados as metas pré-estabelecidas).

O desempenho mensurado deve ser avaliado por gestores e pode servir para integrar diferentes áreas de uma empresa, pois dados de um setor podem ter importância para outro (CALLADO; CALLADO; ALMEIDA, 2007).

Os indicadores também podem ser divididos em cunho financeiro e não-financeiro (CALLADO; CALLADO; ALMEIDA, 2007), e qualitativos quando tratam de atributos e quantitativos ao dizerem respeito a variáveis (HARRINGTON, 1993).

Indicadores qualitativos mostram-se mais fáceis de serem elaborados e tem análise mais abrangente, por outro lado dependem de poder de argumentação do avaliador e aspectos não relevantes podem tornar o texto longo e confuso. Já os quantitativos tem sua interpretação mais clara e sua conclusão mais direta quanto a meta estabelecida, apesar de terem coleta de dados geralmente com alto custo (FRANCISCHINI; FRANCISCHINI, 2017).

Outra característica que pode dividir indicadores é a de serem reativos ou proativos, sendo que os reativos são utilizados para o controle de uma atividade em que se busca uma restauração de desempenho ou alcançar um novo patamar ainda não atingido. Já os proativos tem o intuito de utilizar os indicadores para propor novas ações futuras (MARTINS; COSTA NETO, 1998).

Dentro de uma organização, os indicadores podem estar presentes desde a parte financeira até a ambiental, passando por recursos humanos, projetos sistemas de informação, marketing, produção e armazenamento (CALDEIRA, 2016).

Dentro do âmbito de indicadores, o termo "desempenho" surge juntamente com

outros, formando conceitos pertinentes ao assunto. Neely, Gregory e Platts (1995, p. 80-81) ressaltam o significado de termos:

- Mensuração de desempenho - processo de quantificação da eficiência e eficácia das ações;
- Medida de desempenho - métrica utilizada para quantificar as ações;
- Sistema de medição de desempenho - conjunto de métricas utilizadas para quantificar a eficiência e eficácia das ações.

2.3.1 Sistemas de medição de desempenho

Para dar andamento em suas ações estratégicas, as organizações devem possuir um sistema de medição de desempenho que se utilize de indicadores que auxiliem a tomada de decisão (NAKAMURA; MINETA, 2001).

Há algumas formas de se avaliar o desempenho mensurado: de maneira individual, em conjuntos de medidas, chamados de sistema de medição de desempenho, e verificando a relação entre o sistema de medição e o ambiente em que está incluído (NEELY; GREGORY; PLATTS, 1995).

Os sistemas de medição de desempenho (SMD) devem ser concebidos para ajudar as empresas a alcançarem seus objetivos, mostrando causas de problemas e corrigindo rumos (FRANCISCHINI; FRANCISCHINI, 2017).

É tarefa dos gestores buscar a forma de medir desempenho organizacional, focando na mensuração de eficiência, eficácia, efetividade, qualidade, produtividade, inovação, lucratividade entre outros (PETRI, 2005).

Porém, definir o que deve ser medido pelas empresas é tarefa complicada. Cada setor pode contar com seu conjunto específico de indicadores de desempenho (CALLADO; CALLADO, 2011a).

Devido a este fato da especificidade de cada setor, há também a complicação em gerar indicadores padrões, e conseqüentemente muitas vezes a impossibilidade de comparação entre empresas e processos (NUINTIN; NAKAO, 2010).

Na construção de sistemas de medição, os gestores devem atentar para não cometer erros como excesso de indicadores, indicadores sem utilização e metas, definição de indicadores antes de objetivos, falta de compreensão do que foi medido, periodicidade baixa da disponibilização dos resultados, coleta de dados equivocada e a análise dos indicadores não gerar ações de melhora dentro da organização (FRANCISCHINI; FRANCISCHINI, 2017).

Para se montar um sistema de medição com indicadores, algumas questões devem como "quais as medidas de desempenho serão utilizadas?", "para que serão utilizadas?" e "quanto custam e quais os benefícios?" devem ser respondidas (NEELY; GREGORY; PLATTS, 1995).

A literatura apresenta diversos sistemas de medição de desempenho, alguns deles estão indicados no QUADRO 3

QUADRO 3 – SISTEMAS DE MEDIÇÃO DE DESEMPENHO

Modelo	Fonte
<i>Balanced Scorecard</i> (BSC)	Kaplan e Norton (1992)
<i>Integrated and Dynamic Performance Measurement System</i> (IDPMSa)	Bititci, Carrie e McDevitt (1997) Bititci, Carrie e Turner (1998) Bititci, Turner e Begemann (2000)
<i>Integrated and Dynamic Performance Measurement System</i> (IDPMSb)	Ghalayini e Noble (1996); Ghalayini, Noble e Crowe (1997)
Método de Buchele	Buchele (1967)
Método de Martindell	Martindell (1950)
Modelo Quantum de Medição de Desempenho (MQMD)	Hronec (1994)
<i>Performance Measure Questionnaire</i> (PMQ)	Dixon, Jr. e Vollmann (1990)
<i>Performance Prism</i> (PP)	Neely, Adams e Crowe (2001)
<i>Operating Profit Through Investment Management</i> (OPTIM)	Sullivan (1986)
Sete Critérios de Desempenho (SCD)	Sink e Tuttle (1993)
Sistema de Medição de Desempenho Global	FPNQ (2001)
<i>Strategic Measurement and Reporting Technique</i> (SMART)	Lynch e Cross (1991)
<i>Tableau de Bord</i>	Engenheiros Franceses

Fonte: Figueiredo (2003); Hournaux Junior (2005); Nutini (2015)

Cada organização pode-se utilizar de algum dos modelos, dependendo da necessidade. O *Balanced Scorecard*, modelo mais citado na literatura, foi concebido por Kaplan e Norton (1992) para ser utilizado com indicadores de quatro perspectivas: financeira, de clientes, processos internos e aprendizado e crescimento.

O método de Martindell (1950) busca fazer uma análise completa de uma empresa, contando com dez itens e cada um com sua respectiva pontuação máxima. Já o modelo de Buchele (1967) trabalha com a avaliação do desempenho sob as perspectivas: financeira, de processos administrativos e de departamentos.

O *Strategic Measurement and Reporting Technique* (SMART) trabalha com uma pirâmide que compreende três níveis de objetivos e medidas: visão corporativa estratégica (1º nível), mercado e finanças (2º nível) e operações (3º nível). O critério de resultados do Sistema de Medição de Desempenho Global do (FUNDAÇÃO NACIONAL, 2007) fala sobre o uso de indicadores em quesitos como liderança, estratégias e planos,

clientes e sociedade; informação e conhecimento; pessoas, processos e resultados da organização (FIGUEIREDO, 2003).

O *Tableau de Bord* (TB) mostra-se como um gerador de informação para tomada de decisão através da disponibilização de indicadores considerados importantes para gestores públicos ou empresariais (FAGUNDES et al., 2007).

O TB trabalha com indicadores financeiros e não-financeiros, utilizando-se de gráficos para disponibilizar a informação em tempo real (LÓPEZ, 1999 apud QUESADO; GUZMÁN; RODRIGUES, 2012).

2.3.2 Indicadores de Desempenho e Agronegócio

A competitividade no agronegócio do século XXI fez com que ações fossem tomadas pelas organizações com intuito de destacá-las no mercado. Uma das ações é a aprimorar processos produtivos utilizando indicadores (ALBANO; WOHLLENBERG; GARCIA, 2014).

Para alcançar maior eficiência, sistemas de gestão e controle que forneçam informação de desempenho para aperfeiçoamento das atividades mostram-se como valioso aliado dos gestores (CALLADO; CALLADO; ALMEIDA, 2007).

Assim como nos demais ramos, definir o que deve ser medido e quais as métricas utilizadas é tarefa complicada. A definição de indicadores faz parte de uma sequência lógica onde se encontram desenvolvimento e implementação de um sistema de mensuração e avaliação do desempenho, tendo como objetivo gerar medidas que sejam os objetivos e metas da organização (CALLADO; CALLADO; ALMEIDA, 2007).

Apesar de ser um setor específico, com suas próprias necessidades, o setor do agronegócio tem características semelhantes a outros ramos, e por consequência indicadores comuns (CALLADO; CALLADO; MACHADO, 2007).

Medir produtividade pode auxiliar no melhor desempenho organizacional, pois os indicadores podem mostrar padrões, e se metas estabelecidas foram alcançadas (PAULA; ICHIKAWA, 2002).

Dentro do agronegócio, é possível tanto analisar indicadores de uma organização de forma individual quanto uma cadeia inteira (CALLADO; CALLADO, 2011a).

A mensuração do desempenho se faz presente nas três fases da cadeia agroindustrial (antes, dentro e depois da porteira), sendo utilizada nas atividades agrícolas, pecuárias, de transformação e complementares (CALLADO; CALLADO, 2011a).

O agronegócio, em cada uma de suas fases, pode fazer o uso de indicadores diferentes que se encaixam dentro das especificidades de cada uma delas. Desta forma, o agronegócio pode contar com indicadores de eficácia, eficiência, qualidade,

produtividade, inovação, custo e lucratividade (CALLADO; CALLADO; MACHADO, 2007).

A literatura mostra que fase dentro da porteira conta com indicadores de desempenho financeiro (QUADRO 4) e não-financeiro (QUADRO 5).

QUADRO 4 – INDICADORES FINANCEIROS DO AGRONEGÓCIO

Indicador (unidade de medida)	Descrição	Fonte
Arrecadação bruta (reais)	Arrecadação total em reais de um produto, segmento, fazenda etc.	Gasques et al. (2004) Melo e Vilela (2004) Callado, Callado e Almeida (2007) Alves et al. (2009) Chaviano et al. (2018)
Arrecadação líquida (reais)	Arrecadação em reais de um produto, segmento, fazenda etc descontando as despesas	Alves et al. (2009) Pereira et al. (2016) Chaviano et al. (2018)
Fatia de mercado (porcentagem)	Porcentagem correspondente a fatia de mercado de um país, cidade, empresa etc	Gasques et al. (2004) Contini et al. (2006) Callado, Callado e Almeida (2007)
Despesas (reais)	Total de despesas de um acerca de um produto, segmento, empresa etc	Kreuz, Souza e Clemente (2007) Alves et al. (2009) Pereira et al. (2016) Chaviano et al. (2018)
Preço do produto (reais)	Valor unitário de venda do produto	Callado, Callado e Almeida (2007) Alves et al. (2009) Pereira et al. (2016)
Custo unitário (reais)	Custo para se produzir uma unidade de um produto	Callado, Callado e Almeida (2007) Alves et al. (2009) Pereira et al. (2016)
Magem de lucro (porcentagem)	Taxa de retorno obtida com a comercialização de um produto	Callado, Callado e Almeida (2007)

Fonte: o Autor (2018)

Quanto aos não-financeiros, indicadores de produtividade não-financeiros (QUADRO 5 se mostram importantes dentro da porteira. Paula e Ichikawa (2002) cita que a produtividade está ligada com o total de algo que foi produzido.

Para corroborar com Callado, Callado e Machado (2007), que afirmaram que indicadores utilizados no agronegócio são comuns a outras áreas, verificou-se que os indicadores financeiros encontrados em trabalhos científicos sobre o agronegócio também figuram em outros trabalhos como por exemplo Santos e Miranda (2015), que

abordam a arrecadação líquida num ramo de vendas, Taffarel e Garcias Paulo Mello (2009) que citam arrecadação bruta e líquida de empresas alimentícias e mais recentemente Scheren, Wernke e Zanin (2018), que abordam despesas, arrecadação bruta e líquida no contexto da saúde.

QUADRO 5 – INDICADORES NÃO-FINANCEIROS DO AGRONEGÓCIO

Indicador (unidade de medida)	Descrição	Fonte
Área colhida (mil hectares)	Área colhida total de alguma cultura	Contini et al. (2006) Baccarin, Gerbara e Borges Junior (2011) Maia et al. (2015)
Área plantada (mil hectares)	Total da área plantada de alguma cultura	Gasques et al. (2004) Melo e Vilela (2004) Contini et al. (2006) Sousa, Bonjour e Figueiredo (2006)
Produção (mil litros)	Quantidade litros produzidos a partir de uma fonte	Alves et al. (2009) Baccarin, Gerbara e Borges Junior (2011) Pereira et al. (2016) Chaviano et al. (2018)
Produção (mil toneladas)	Aponta o total de produção de um produto em toneladas	Gasques et al. (2004) Melo e Vilela (2004) Contini et al. (2006) Sousa, Bonjour e Figueiredo (2006) Alves et al. (2009) Baccarin, Gerbara e Borges Junior (2011) Maia et al. (2015) Chaviano et al. (2018)
Produtividade (kg/hectares)	Razão para obter a média de quilogramas por hectare	Gasques et al. (2004) Melo e Vilela (2004) Contini et al. (2006) Sousa, Bonjour e Figueiredo (2006) Maia et al. (2015)
Rebanho (mil cabeças)	Quantidade de animais	Alves et al. (2009) Pereira et al. (2016) Chaviano et al. (2018)
Volume Processado (toneladas)	Quantidade de um determinado produto que foi processado	Melo e Vilela (2004)

Fonte: o Autor (2018)

Os indicadores citados são indicados nos trabalhos acompanhados, geralmente,

de períodos temporais. A medição dos indicadores pode ser feita de forma diária, semanal, mensal, por safra, anual e etc.

Alguns indicadores podem ser utilizados tanto na pecuária quanto na agricultura. A produção em toneladas pode ser, por exemplo, de tomate, como no trabalho de Melo e Vilela (2004) ou de carne bovina, como no estudo de Chaviano et al. (2018). Já a produção em litros pode ser de leite, produto agropecuário, como indicada por Pereira et al. (2016), e também de produtos agrícolas, como o álcool vindo da cana-de-açúcar no citado de Baccarin, Gebara e Borges Junior (2011).

2.4 FONTES DE DADOS E INFORMAÇÕES

A concorrência nos negócios faz com que gestores se tornem usuários da informação, que por sua vez devem ser precisas relevantes e confiáveis. Este panorama obriga as organizações a compreenderem aspectos informacionais, e isso inclui entender critérios para a escolha de fontes de informações (PEREIRA, 2016).

As fontes de dados dentro do âmbito organizacional são obtidas com o intuito de agregar elementos específicos de um determinado ambiente para a empresa, visando vantagem competitiva (BARBOSA, 2006). Estas fontes podem ser classificadas como interna ou externas (KOBRIIN et al., 1980).

O estudo de Carmo e Pontes (1999) aponta como exemplos de fontes externas clientes, fornecedores, concorrentes, consultores, amigos, universidades, bases de dados, livros e jornais, enquanto as internas apontadas são basicamente relatórios de diversos tipos como contábeis, de vendas, produção e recursos humanos.

Além da classificação como interna ou externa, Degent (1986) indica como possíveis fontes de dados às organizações: concorrentes, governo, fornecedores, clientes, associações profissionais, associações de classe, empregados e outras fontes como imprensa e consultores de marketing por exemplo.

De acordo com Montalli (1991), as fontes de informação podem ser:

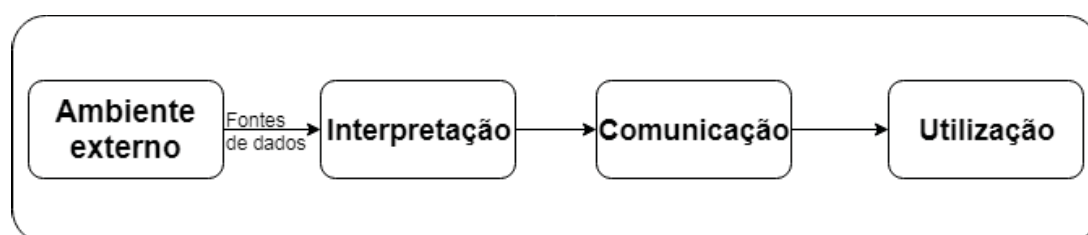
- Científicas - livros, periódicos, anais de congresso, artigos de revisão, conferências e bases de dados;
- Técnicas - manuais, regulamentos, patentes, legislações;
- Negócios - relatórios anuais, produtos, serviços e relatórios de pesquisa, dados sobre mercados, manuais, revistas de empresas, publicações estatísticas.

Os dados contidos dentro das fontes também podem ser classificados quanto a sua estrutura. Dados chamados de estruturados são aqueles armazenados em bancos

de dados, que por sua vez contam com formato rígido, ou seja, todos tem, obrigatoriamente, a mesma estrutura. Os semiestruturados contemplam uma estrutura, mas não é necessário que todos os dados contenham a mesma estrutura, e os não-estruturados tem como exemplo páginas HTML, onde a informação embutida se encontra sem uma estrutura formal (ELMASRI; NAVATHE, 2005).

Para que os dados contidos em uma fonte tornem-se utilizáveis para a tomada de decisão, eles são submetidos a um ciclo (Figura 13).

Figura 13 – CICLO DA UTILIZAÇÃO DE FONTES EXTERNAS DE DADOS



Fonte: Barbosa (1997)

Segundo Barbosa (1997), o ambiente externo é composto por clientes, concorrentes, economia, regulamentações entre outros, que fornecem dados que devem ser interpretados pela organização. A parte de comunicação diz respeito a formação do consenso acerca dos dados e compartilhamento das informações posteriormente. A finalização do ciclo ocorre com utilização da informação para o planejamento estratégico e a tomada de decisão.

De acordo com Auster e Choo (2004), cabe ao gestores utilizar seletivamente o que está contido dentro da fonte externas de dados para formular estratégias e tomar decisões.

Ao procurarem informações sobre certo ramo de negócio, gestores por vezes se deparam com diversas fontes externas de dados, tendo como consequência a dificuldade na seleção de uma delas, pois torna-se complexo diagnosticar quais contem o que realmente se precisa dentro da organização (BARBOSA, 1997) (AUSTER; CHOO, 2004).

Em seu trabalho, Auster e Choo (2004) levanta pontos em que o gestor deve refletir antes realizar a seleção da fonte de dados:

- Acessibilidade - verificar qual o tamanho do esforço para se localizar e obter os dados da fonte;
- Qualidade - diz respeito a relevância e a confiabilidade. A relevância está ligada

ao quanto o dados são importantes para a organização, e a confiabilidade no quanto o gestor confia pessoalmente nos dados;

- Uso da fonte - levantar com que frequência a informação proveniente da fonte de dados será utilizada;

2.4.1 Extração e tratamento de dados

Para que dados de fontes externas possam ser utilizados, é necessário a extração, tratamento e carregamento dos mesmos. Esses processos realizados de forma automatizados são chamados de *Extract, Transform, Load* (ETL) (INMON, 2005).

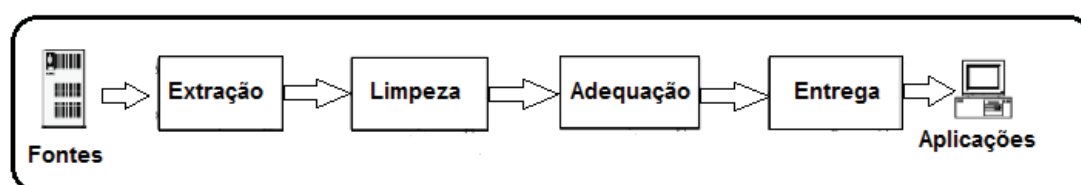
A ETL é utilizada, principalmente, em *data warehouses*, realizando periodicamente a extração e combinação de dados de diversas fontes em uma única(ECKERSON; WHITE, 2003).

O processo de ETL, de acordo com Kimball e Caserta (2004), tem como objetivos: extrair, reforçar a qualidade e ajustar dados de diferentes fontes para que os mesmos possam ser juntados e entregues em um formato que possibilite a criação de soluções para a tomada de decisão. Um bom sistema de ETL deve ser capaz de se comunicar com diferentes fontes de dados em diferentes formatos (FERREIRA et al., 2010); de acordo com os autores, é função da ETL:

- Remoção e correção de dados;
- Fornecer medidas da confiança dos dados;
- Trata dados de diferentes fontes para que possam ser agrupados;
- Estruturação dos dados para serem utilizados por usuários finais através de ferramentas

Para que as funções da ETL sejam realizadas, os dados são submetidos a um ciclo de vida de quatro estágios (Figura 14).

Figura 14 – ESTÁGIOS DOS DADOS NA ETL



Fonte: Kimball e Caserta (2004)

O procedimento de extração corresponde a coleta de dados de fontes externas. Cada fonte possui seu próprio conjunto de características que precisam ser gerenciadas para uma extração de sucesso (EL-SAPPAGH SHAKER H. ALI ANDHENDAWI; BASTAWISSY, 2011).

No campo científico, a extração de dados de fontes pode ser feita de maneira manual ou de forma automática conforme indicado pela metodologia ETL, utilizando-se de ferramentas geradas especificamente para esta finalidade, como realizado por Magalhães et al. (2014) e Furtado et al. (2015).

A limpeza diz respeito a identificação e correção de erros e omissões nos dados, enquanto a adequação cuida da resolução de conflitos na rotulagem dos dados para que possam ser utilizados juntos (KIMBALL; CASERTA, 2004).

Na metodologia científica, muitas vezes dados adquiridos de fontes externas necessitam passar por procedimentos de transformação para que possam ser utilizados. De acordo com Marconi e Lakatos (2007), os dados devem ser submetidos a:

- Seleção - trata-se do exame do que foi obtido, a verificação dos dados com intuito de buscar possíveis erros, informações confusas, distorcidas e incompletas;
- Codificação - processo de categorização e transformação dos dados em símbolos, deixando a possibilidade de tabelá-los e contá-los;
- Tabulação - disposição dos dados em tabelas, com objetivo de facilitar a visualização da relação entre eles

Uma das formas de transformação de dados consiste em organizá-los com luz no conceito chamado *tidy data*. O *tidy data* é classificado por (WICKHAM, 2014) como um padrão de organização de dados que visa simplificar o desenvolvimento de ferramentas de análise de dados. De acordo com o autor, *tidy data* segue as seguintes regras:

- Cada variável medida deve estar em uma coluna;
- Cada observação desta variável deve estar em uma linha;
- Uma tabela para cada tipo de variável

Wickham (2014) ainda indica cinco tipos de problemas comuns em fontes de dados que devem ser sanados para aplicação das regras:

- Cabeçalhos são valores e não descrição de variáveis;

- Mais de uma variável armazenada em uma coluna;
- Variáveis armazenadas em linhas e colunas;
- Variáveis diferentes armazenadas na mesma tabela;
- Uma única variável aparece em várias tabelas

Em seu trabalho, Wickham e Grolemund (2017) exemplificam algumas transformações que devem ser feitas em dados para aplicação de *tidy data* (Figura 15).

Figura 15 – TRANSFORMAÇÃO DE DADOS POR *TIDY DATA*

country	year	cases
Afghanistan	1999	745
Afghanistan	2000	2666
Brazil	1999	37737
Brazil	2000	80488
China	1999	212258
China	2000	213766

country	1999	2000
Afghanistan	745	2666
Brazil	37737	80488
China	212258	213766

table4

Fonte: Wickham e Grolemund (2017)

A figura apresenta a tabela da esquerda como originária da transformação da tabela direita seguindo conceitos de *tidy data*. A tabela de origem conta com problemas como: variável estar contida em mais de uma coluna e cabeçalhos serem valores e não descrições. Para aplicar o conceito de *tidy data*, os autores trazem como resultado uma tabela onde cada caso está contido em uma linha, e as colunas que eram valores (1999 e 2000) passaram a se chamar "anos".

Para alcançar cumprir as regras determinadas, as variáveis devem ser tratadas uma a uma, passando por: filtragem, onde são agrupadas em subconjuntos ou removidas; transformação, classificada como procedimento de modificação das variáveis; agregação, realizando o agrupamento de valores comuns e por último a ordenação dos dados (WICKHAM, 2014).

O último estágio de ETL diz respeito ao carregamento do arquivo com os dados que foram extraídos, limpos e agrupados, com o intuito de servir as aplicações e usuários finais (EL-SAPPAGH SHAKER H. ALI ANDHENDAWI; BASTAWISSY, 2011).

A integralização de dados de diferentes fontes requer que os procedimentos de ETL sejam personalizados para cada uma delas, sobretudo quando se tratam de dados

oriundos de fontes relacionais, onde o processo ETL deverá impor a regra de negócio (KIMBALL; CASERTA, 2004).

Assim como a extração, a transformação e o carregamento dos dados também podem ser realizados de forma manual ou automatizada via alguma ferramenta disponível ou *script* específico que realize as ações de ETL, trabalhando com diversos formatos de arquivos com dados.

2.4.2 Estruturas de Dados

Além de ser comum os dados serem originários de várias fontes, é normal também que se sejam de diferentes formatos (KIMBALL; CASERTA, 2004). Arquivos com diferentes estruturas disponibilizam dados e possibilitam a mineração dos mesmos, entre eles encontram-se:

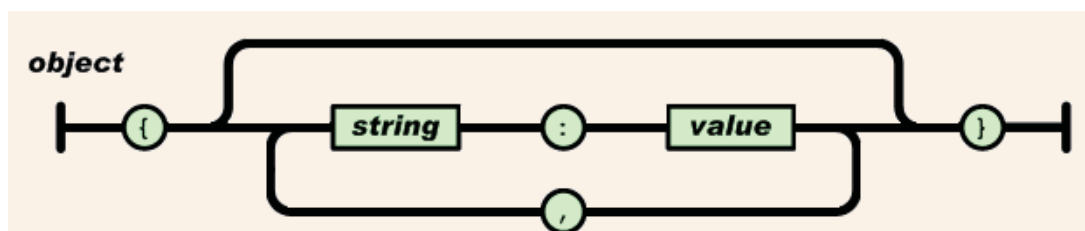
- **Arquivo de Texto (.txt)** - são arquivos simples, contendo textos dentro de seu conteúdo. O texto incluso dentro do arquivo segue o padrão de codificação *American Standard Code for Information Interchange* (ASCII).
- **Comma-Separated Values (.csv)** - arquivos separados por vírgula são utilizados em softwares que trabalham com planilhas, como por exemplo Microsoft Excel, LibreOffice Calc e a linguagem de programação R.

Apesar de ser um amplamente utilizado, este tipo de arquivo não é formalmente documentado, possibilitando diversas especificações e implementações, permitindo, por exemplo, que o usuário determine o delimitador de campos, quebra de linhas e cabeçalhos (SHAFRANOVICH, 2005).

- **Data Base (.db)** - documentos de formato .db contemplam bases de dados estruturados em tabelas podem ser utilizados em softwares como Microsoft Excel e Microsoft Access e transformado em arquivo CSV.
- **Extensible Markup Language (.xml)** - assim como o *HyperText Markup Language* (HTML), o XML é uma linguagem de marcação de texto, desenvolvida em 1996, derivada do Standard Generalized Markup Language (SGML) (W3C, 2006). O XML permite que o usuário crie suas próprias marcações, o que acaba por fornecer a facilidade de entendimento do código escrito. Conforme a (W3C, 2006), documentos deste tipo tem como características: serem legíveis e claros, feitos rapidamente e facilmente, formais, concisos e utilizáveis na internet.
- **JavaScript Object Notation (json)** - formato de compactação de para troca de dados. O json traz a facilidade para escrita e leitura de humanos e interpretação e construção em máquinas (JSON.ORG, 2018).

A passagem de dados em json é feita através de objetos (Figura 16), que contemplam conjuntos desordenados de pares de valor e nome (JSON.ORG, 2018).

Figura 16 – FORMATO DO OBJETO JSON DE TROCA DE DADOS



Fonte: Json.org (2018)

Os dados de objetos enviados via json são separados através de vírgulas. Além da passagem por objetos, o json permite também o envio de vetores de dados, sendo sep

- **Portable Document Format (.pdf)** - o tipo de arquivo criado em 1991 pelo cofundador da *Adobe Systems Incorporated* e mantido atualmente pela *Organization for Standardization* (ISO) tem como objetivo principal facilitar o compartilhamento de documentos compatibilizando a visualização de documentos independente de software, hardware ou sistema operacional (ADOBE, 2018).

Arquivos .pdf possibilitam a inclusão de links dentro de seus documentos (ADOBE, 1997). Campos de formulário, áudio, vídeo, botões, assinaturas digitais e indicação de senha para proteção do conteúdo são outras possibilidades de inclusão em arquivos deste tipo (ADOBE, 2018).

- **Structured Query Language (.sql)** - arquivos que contem a extensão .sql contem em si um conjunto de instruções, em linguagem SQL, para criação de um banco de dados relacional, ou ao menos instruções para ações como: inclusões, exclusões e alterações de registros ou estruturas de um banco já existente.
- **.xls e .xlsx** - .xls e .xlsx são arquivos de planilhas que, assim como arquivos .csv, funcionam com softwares como o Microsoft Excel e LibreOffice Calc e a linguagem de programação R.

Além da inclusão de dados, arquivos XLS e XLSX permitem a inclusão de gráficos, fórmulas matemáticas e estatísticas. Documentos XLSX passaram a ser utilizados a partir da versão 2007 do software Microsoft Excel.

2.4.3 ETL e Linguagens de Programação

As linguagens de programação e suas bibliotecas permitem a construção de ferramentas que, através de seus algoritmos, atuam nas três etapas de ETL.

Uma das formas de se cumprir a primeira etapa de ETL é através da técnica chamada *web scraping*, onde há a extração dos dados diretamente de endereços eletrônicos da internet através de programas de computador (HADDAWAY, 2015); (VARGIU; URRU, 2013).

É possível realizar *web scraping* em diversas linguagens de programação diferentes como Java (MITCHELL, 2013), PHP (TURLAND, 2010), *Python* (MITCHELL, 2015) e R (MUNZERT et al., 2014).

A linguagem *Python*, caracterizada como de alto nível, interpretada e orientada a objetos (PYTHON.ORG, 2018) é um exemplo de linguagem que permite a construção de algoritmos que compõem as três etapas de ETL. Sua alta difusão no meio científico justifica-se pelo fato de suas bibliotecas oferecerem ferramentas para análise e manipulação de dados (NELLI, 2015).

Exemplo disso é a ferramenta ScriptLattes, construída em *Python* e utilizada na extração de dados de pesquisadores na plataforma Lattes do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) conforme indicado por Mena-Chalco e Cesar Junior (2013).

Bibliotecas como a *BeautifulSoup*, utilizada na linguagem *Python*, permitem o trabalho de mineração de dados em páginas HTML e em arquivos XML, permitindo a geração de algoritmos específicos de extração de dados de arquivos destes tipos (NAIR, 2014).

Com relação a parte de transformação dos dados, *Python* conta com bibliotecas para este tipo de trabalho. Exemplo disso é a biblioteca Panda, utilizada para transformar dados dispostos em estruturas de linhas e colunas, conhecidas como *data frames* (MCKINNEY, 2012).

A parte de carregamento dos dados pode ser feita em *Python* utilizando-se da biblioteca *Matplotlib*, utilizada para ler dados de *data frames* e plotar gráficos de resultados da leitura (MCKINNEY, 2012). Nelli (2015) aponta qualidades da biblioteca *Matplotlib*: fácil uso, controle sobre os elementos gráficos e possibilidade de exportação dos gráficos em vários formatos.

A linguagem de programação R, derivada da linguagem S, é muito utilizada para computação estatística e gráfica sendo *software* livre (R CORE TEAM, 2018); (R, 2018); (VERZANI, 2011). A primeira da etapa de ETL pode ser feita na linguagem R de maneiras diferentes, dependendo da necessidade da situação.

Para realização de *web scraping*, a linguagem R conta com bibliotecas como RCrawler, que possibilita o acesso e *download* do conteúdo de páginas web via comando Rcrawler (KHALIL; FAKIR, 2017). Outra biblioteca que realiza trabalho semelhante é a XML2, que possibilita ler o código-fonte de páginas da internet.

O uso de bibliotecas de leitura e *download* de códigos-fonte pode ser combinado com a biblioteca Rvest, que possibilita extrair dados contidos dentro dos códigos-fonte através do mapeamento de tags contidas nos arquivos (KHALIL; FAKIR, 2017); (BOEHMKE, 2016).

Para o caso de disponibilização de dados em JSON, a linguagem R conta com a biblioteca Rjson, que permite a captura do JSON diretamente da internet para que os dados possam ser tratados posteriormente (COUTURE-BEIL, 2018).

Muitos portais eletrônicos disponibilizam, através de *links* para download, seus dados em arquivos de planilhas, PDF e etc. Para isso, a biblioteca XML permite a fácil identificação destes *links* para download dos arquivos (MUNZERT et al., 2014, p. 223). A biblioteca Readxl permite a leitura de arquivos de planilhas em formatos .XLSX, .XLS e .CSV, para possam ser carregados para serem submetidos ao processo de transformação de dados, segunda etapa de ETL (BOEHMKE, 2016).

Diversas funções possibilitam a manipulação dos dados inclusos dentro de um arquivo, como por exemplo *substr* para utilização de substrings, *numeric* para conversão de dados tipo numérico, *colnames* para criação e alteração de colunas de dados dentro de planilhas utilizadas no projeto.

Além das funções nativas da linguagem R, a biblioteca Dplyr possui funções como *select*, *filter*, *group_by* e *mutate* para selecionar colunas, filtrar e agrupar linhas por atributo e alterar colunas (WICKHAM et al., 2015). Estas funções são baseadas em ações análogas da linguagem SQL, mas podem ser utilizadas em conjunto com as diversas funcionalidades para manipulação de dados da linguagem R.

Funções citadas para manipulação dos dados permitem a uniformização dos dados. A linguagem R permite a criação de tabelas de dados chamadas de *data frames*, utilizadas para guardar dados tratados (BIVAND; PEBESMA; GÓMEZ-RUBIO, 2008).

Para a parte de carregamento dos dados, a biblioteca ggplot2 possibilita a plotagem de diversos tipos de gráficos, utilizando-se de dados de planilhas. A biblioteca permite, por exemplo, a criação de gráfico de barras, gráfico de linhas, boxplots e vários outros tipos de representações (WICKHAM, 2009); (MOON, 2017).

A biblioteca Shiny possui suporte para a visualização de gráficos nativos da linguagem R, bem como de outras bibliotecas gráficas como a ggplot2, em navegadores da internet (WICKHAM; GROLEMUND, 2017). Além da visualização, a *Shiny* permite a criação de interfaces interativas que reagem a entradas do usuário, trazendo a

oportunidade de interação do usuário com a massa de dados da ferramenta (BEELEY, 2013); (WICKHAM; GROLEMUND, 2017).

Assim como Python e R, outras linguagens como PHP permitem também a retirada e limpeza de dados vindos de fontes de *internet* (TURLAND, 2010); (WARD, 2013); (OIKAWA, 2017).

2.4.4 Metadados

O uso de sistemas de informação e o consequente acúmulo de dados, traz a necessidade da utilização de metadados, classificados por XXX como dados sobre dados. Esta definição significa que metadados trazem consigo informações acerca de um determinado recurso, possibilitando ao usuário, segundo Riley (2017), encontrar itens de seu interesse, registrar e compartilhar informações com outras pessoas.

A necessidade de organizar informações é que deu origem aos metadados (JUNIOR; OLIVEIRA, 2017), que se popularizaram a partir da década de 90 a partir da criação de diferentes modelos para atender a necessidade de distintos tipos de informação (BRAMBILLA; CARRO; FELÍCIO, 2016).

De acordo com Riley (2017), os metadados podem ser divididos em três tipos:

- Metadados Descritivos - utilizados para encontrar e descrever um recurso;
- Metadados Administrativos - subdivide-se em metadados técnicos, de preservação e de direitos, utilizados para decodificação e renderização de arquivos, gerenciamento de termos técnicos de arquivos e associação de direitos autorais ao recurso;
- Metadados Estruturais - indicam como ocorre o relacionamento de partes do recurso entre si.

Apesar de não haver uma linguagem padrão para apresentação de metadados, os formatos XML e RDF são os mais utilizados para esta finalidade (SICILIA; LYTRAS, 2009). Assim como na parte de linguagens, os metadados podem ser construídos de acordo com diferentes formas, tendo como um dos destaques o esquema Dublin Core da Dublin Core Metadata Initiative (DCMI), indicado em trabalhos como de Souza, Vendrusculo e Melo (2000); Nilsson et al. (2009); e mais recentemente por Bentancourt e Rocha (2012) e Brambilla, Carro e Felício (2016).

O Dublin Core tem como objetivo facilitar a maneira de descrever recursos eletrônicos, e, conseqüentemente, tornar estes recursos mais visíveis a motores de busca (SOUZA; VENDRUSCULO; MELO, 2000). De acordo com Weibel et al. (1998),

o Dublin Core trabalha com 15 elementos relacionados ao: conteúdo, propriedade intelectual e características formais do recurso descrito (QUADRO 6).

QUADRO 6 – ELEMENTOS DO DUBLIN CORE

Elementos de conteúdo	Elementos de propriedade intelectual	Elementos de características formais
Título	Criador	Data
Assunto	Colaborador	Formato
Relações	Editor	Identificador
Fonte	Direitos	Idioma
Cobertura		
Tipo		
Descrição		

Fonte: Weibel et al. (1998)

De acordo com Bentancourt e Rocha (2012), os 15 elementos podem ser assim descritos:

- **Título** - Nome pelo qual o recurso é conhecido;
- **Criador** - Identificação do responsável pela criação do recurso;
- **Assunto** - Diz respeito ao conteúdo do recurso. Geralmente expresso em palavras-chave ou códigos de classificação;
- **Descrição** - Tem como objetivo descrever o conteúdo, geralmente se utilizando de resumos, sumários ou textos livres;
- **Editor** - Responsável por tornar o recurso disponível, seu uso não é recomendado no caso do editor e criador serem a mesma pessoa ou organização;
- **Colaborador** - Neste tópico são representados aqueles que contribuíram com o conteúdo do recurso;
- **Data** - Representa alguma data do ciclo de vida do documento. Informação utilizada em formato AAAA-MM-DD (ano-mês-dia).
- **Tipo** - Indica a natureza ou gênero do conteúdo. Geralmente, utiliza-se de um vocabulário genérico da Dublin Core Metadata Initiative (DCMI);
- **Formato** - Manifestação física ou digital do recurso. Além do tipo da mídia, dimensões de tamanho ou duração, também pode indicar softwares e equipamentos para melhor uso do recurso;

- **Identificador** - Referência ao um recurso em um dado contexto, por exemplo: Digital Object Identifier (DOI), International Standard Book Number (ISBN) e etc;
- **Fonte** - Indica referências ao qual o recurso é derivado;
- **Idioma** - Linguagem do conteúdo intelectual do recurso; geralmente se utiliza das especificações RFC 3066 e ISO 639 para identificação de idiomas;
- **Relações** - Referência a um determinado recurso relacionado, geralmente feito através de texto ou um número que identifique a referência;
- **Cobertura** - Identificação espacial ou temporal do conteúdo; nome de lugares, coordenadas geográficas ou jurisdição;
- **Direito** - Informações sobre direitos autorais ligados ao recurso disponibilizado.

2.4.5 Fontes de dados do agronegócio

Por contemplar diversos tipos diferentes de indicadores, dados sobre o agronegócio são disponibilizados em diversas fontes, conforme indicado nos trabalhos de Oikawa (2017) e Boteon (2004).

A partir de um levantamento bibliográfico, indicado por Marconi e Lakatos (2007) e Gil (2002), foram levantadas, a partir dos artigos utilizados neste trabalho, organizações que possivelmente disponibilizem fontes de dados de produção do agronegócio (QUADRO 7). Foram considerados órgãos que foram encontrados em mais de uma fonte.

QUADRO 7 – POSSÍVEIS FONTES DE DADOS DO AGRONEGÓCIO BRASILEIRO

Organização	Fontes
Companhia Nacional de Abastecimento (CONAB)	Baccarin, Gebara e Borges Junior (2011), Contini et al. (2006), Maia et al. (2015), Albano, Wohlenberg e Garcia (2014)
Confederação da Agricultura e Pecuária do Brasil (CNA)	Sousa, Bonjour e Figueiredo (2006), Mendonça (2015), Gasques et al. (2004)
Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA)	Sousa, Bonjour e Figueiredo (2006), Elias (2013), Mendonça (2015), Contini e Martha Jr. (2010), Costa (2016), Gasques et al. (2004), Albano, Wohlenberg e Garcia (2014)
Instituto Brasileiro de Geografia e Estatística (IBGE)	Maia et al. (2015), Elesbao (2007), Gasques et al. (2004), Callado, Callado e Almeida (2007), Pereira et al. (2016), Sousa, Bonjour e Figueiredo (2006), Costa (2016), Contini e Martha Jr. (2010)
Ministério da Agricultura, Pecuária e do Abastecimento (MAPA)	Baccarin, Gebara e Borges Junior (2011), Elias (2013), Contini et al. (2006), Mendonça (2015), Gasques et al. (2004), Rocha Junior, Bittencourt e Ribeiro (2015), Ferreira e Camargo (2013), Callado e Soares (2014), Callado, Callado e Machado (2007)
Ministério da Indústria, Comércio Exterior e Serviços	Sousa, Bonjour e Figueiredo (2006), Costa (2016), Contini et al. (2006), Figueiredo, Santos e Lima (2012)
Organização das Nações Unidas para Agricultura e Alimentação (FAO)	Callado, Callado e Almeida (2007), Sousa, Bonjour e Figueiredo (2006), Costa (2016), Abbade (2014), Figueiredo, Santos e Lima (2012), Karnopp e Oliveira (2012)

Fonte: o Autor (2018)

2.5 VISUALIZAÇÃO DA INFORMAÇÃO

Mensagens curtas e objetivas trazem facilidade em seu entendimento. Mensagens textuais longas e elaboradas trazem maior dificuldade na extração e percepção de informações críticas de seu conteúdo (DIAS; CARVALHO, 2007).

Apresentar informações por meio textual é valioso, mas gráficos e diagramas trazem a oportunidade de maior velocidade de transmissão da informação aos gestores, principalmente se tratando de indicadores (FRANCISCHINI; FRANCISCHINI, 2017).

O poder cognitivo da visualização da informação admira estudiosos desde o fim do século XVII e começo do XVIII, como William Playfair, criador dos diagramas de barras e de pizza. No século XX, destaca-se como referência Edward Tufte, estatístico responsável por ligar a visualização da informação com trabalhos quantitativos (MANOVICH, 2011).

A tendência mundial aponta o crescimento exponencial na quantidade de infor-

mações produzidas e veiculadas, além da grande gama de assuntos vinculados as mesmas. Esse panorama gera grande volume de informações provenientes de diversas fontes e formatos (VALIATI, 2008).

A visualização da informação (VI) trabalha na inclusão informacional dos usuários, através de imagens, gráficos, figuras e qualquer outro recurso gráfico capaz de apresentar uma informação que possa ser compreendida de forma mais natural, exigindo menor esforço cognitivo (DIAS; CARVALHO, 2007); (CORREA; VIEIRA, 2013); (NASCIMENTO; FERREIRA, 2005).

As tecnologias computacionais atuais permitem, de acordo com conforme Dias e Carvalho (2007), a produção de sistemas mais complexos, que por sua vez, oferecem informações ricas e precisas no que diz respeito a sua apresentação, caracterizada como agradável, interativa e próxima a realidade. O autor cita como exemplo o aprendizado de pilotos de helicópteros via simulador de vôo, que trazem realidade condizente com a de pilotos habilitados do mundo real.

Campo emergente de pesquisa, a visualização da informação trabalha com a construção de representações visuais de dados abstratos, visando facilitar o entendimento e/ou auxiliar na descoberta de novas informações ali contidas (NASCIMENTO; FERREIRA, 2005). A área obteve ainda mais destaque com a presença dos computadores, que possibilitam uma grande gama de estruturas de interfaces gráficas para apresentação das informações (DIAS; CARVALHO, 2007).

A área da visualização da informação condensa aspectos de computação gráfica, interação humano-computador, cartografia e mineração de dados, com intuito de combiná-los, transformando dados em imagens ou gráficos (CORREA; VIEIRA, 2013). Dois aspectos principais compõem a VI: modelagem estrutural, que tem por objetivo detectar, extrair e simplificar informações ocultas nos dados e representação gráfica da modelagem estrutural realizada (CHEN, 2006).

As técnicas ligadas a visualização da informação procuram representar, através de elementos gráficos, a informação contida em um determinado domínio de aplicação, de modo que estes elementos possam trazer ao ser humano a interpretação e compreensão correta para dedução de novos conhecimentos (FRETAS et al., 2001). A representação da visão geral deve ser feita primeiramente, seguida de zoom e filtros, e por fim detalhes demandados pelo usuário (SHNEIDERMAN, 1996).

A VI tem como objetivo, além de apresentar os dados através de recursos gráficos, fornecer a possibilidade de interação, para que o usuário possa explorar os resultados, com intuito de melhor interpretação e compreensão, resultando numa melhor inferência de novos conhecimentos (VALIATI, 2008).

Vale ressaltar que nem sempre são necessários recursos computacionais para

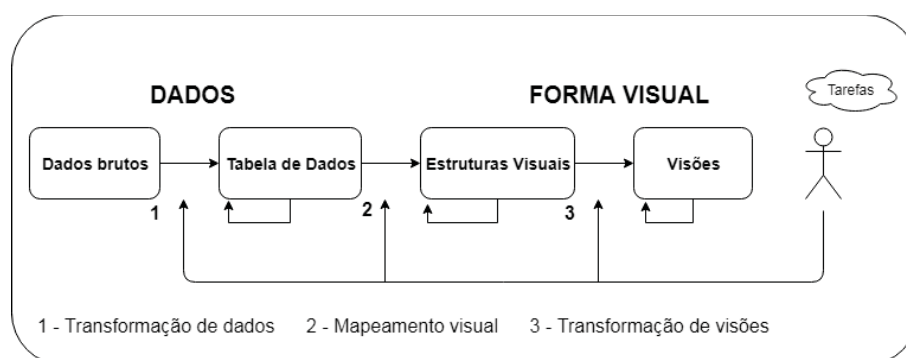
o uso de informações mediante representações gráficas. (DIAS; CARVALHO, 2007) citam como exemplo as placas de sinalizações de trânsito, que são representações gráficas que trazem informações aos motoristas. Os semáforos, através de suas cores, também são caracterizados pelo autor como informações gráficas que auxiliam na tomada de decisão do condutor.

As representações gráficas são caracterizadas por imagens ou figuras utilizadas para representar conjuntos e subconjuntos de dados para serem analisados. Diversos tipos de gráficos como de ponto, de linha, de barra, histogramas geralmente são utilizados, com intuito de demonstrar valores ou relacionamentos entre entidades ou elementos dentro dos dados a serem analisados (FRETAS et al., 2001).

O mais comum é que a visualização da informação demonstre grande quantidade de dados em suas imagens. Tufte (1990) cita como exemplos: mapas, calendários, eletrocardiogramas, fotografias aéreas e até mesmo manuscritos de Galileu.

O processo de transformação de dados em uma representação visual ocorre quase sempre da mesma forma, seguindo o modelo proposto por Card, Mackinlay e Shneiderman (1999) (Figura 17);

Figura 17 – PROCESSO DE TRANSFORMAÇÃO DE DADOS EM REPRESENTAÇÃO GRÁFICA



Fonte: Card, Mackinlay e Shneiderman (1999)

Segundo Card, Mackinlay e Shneiderman (1999), primeiramente os dados brutos são transformados em uma tabela de linhas, colunas e metadados. Após a transformação, a tabela é mapeada para a estrutura visual selecionada para a visualização. Por fim, a estrutura gráfica resultante é mostrada ao usuário, que pode interagir com a mesma.

O uso de técnicas de visualização da informação permite: observar, manipular, filtrar, explorar, pesquisar, navegar, entender e interagir com uma grande massa de dados de forma mais rápida e efetiva no descobrimento de padrões escondidos dentro

os dados (ESTIVALET, 2000).

Segundo Ware (2004), uma das principais vantagens da VI, se bem apresentada, se encontra na possibilidade de se interpretar rapidamente uma grande quantidade de informação. Além disso, o autor aponta cinco outras vantagens no uso da visualização da informação eficiente:

- Compreensão - permite a capacidade de compreensão de uma grande quantidade de informação;
- Percepção - possibilidade de percepção de propriedades emergentes não antecipadas. Padrões percebidos podem servir de base para novas ideias;
- Visualização de problemas - a visualização da informação eficiente permite a rápida percepção de problemas devido ao imediato aparecimento dos mesmos nos dados;
- Compreensão de diferentes recursos - tanto recursos grandes como pequenos podem ser compreendidos
- Formação de hipóteses - auxilia na criação de hipóteses investigativas futuras em cima do que é analisado no presente.

O uso de cores pode também proporcionar melhor disposição de informação ao usuário. Segundo Tufte (1990), trabalhar com cores é algo complexo, deve-se evitar utilizar mais de 20 ou 30 cores, o que pode trazer um retorno negativo. O autor cita que as cores são fundamentais na rotulação, medição, representação ou imitação da realidade e decoração da informação.

Outra questão VI trata da interação do usuário com a informação disposta. Autores como Yi et al. (2007) e Shneiderman (1996) citam exemplos de ações de interação (QUADRO 8).

QUADRO 8 – TIPOS DE INTERAÇÃO DO USUÁRIO NA VISUALIZAÇÃO DA INFORMAÇÃO

Interação - Shneiderman (1996)	Interação - Yi et al. (2007)
Visão Geral - oferecer ao usuário a visão geral de toda a visualização	Seleção - permitir a seleção de itens com intuito de destacá-los. Concede facilidade no rastreamento de um dado durante a análise da informação
Zoom - ampliar itens de interesse do usuário	Exploração - possibilitar a análise de subconjuntos de dados, dispondo e ocultando informações segundo a necessidade do usuário
Extração - possibilitar a extração de subconjuntos de dados	Reconfiguração - fornecer perspectivas de visualização de acordo com o usuário, exemplo: ordenação
Histórico - armazenar ações feitas pelo usuário	Codificação - conceder a alteração da parte visual da representação. Mudanças de cor, formas e tamanho são possibilidades
Detalhes sob demanda - obter detalhes de conjuntos assim que necessário	Abstração/elaboração - proporcionar ajuste do nível de abstração, aumentando ou diminuindo a quantidade de detalhes
Filtragem - permitir a seleção somente de dados convenientes	Filtragem - viabilizar a alteração dos dados mediante condições impostas pelo usuário na base de dados analisada
Relação - exibir o relacionamento entre itens analisados	Conexão - conceder a possibilidade de destaque em relações e associações

Fonte: o Autor (2018)

A combinação de representações gráficas com filtros pode ser importante para a tomada de decisão, fornecendo visões controladas pelo gestor dependendo de suas necessidades.

2.5.1 Representações Gráficas

A representação gráfica da informação extraída de um conjunto de dados depende, primeiramente, das características desse conjunto. Shneiderman (1996) caracteriza sete diferentes tipos de dados:

- 1 dimensão - arquivos textuais, lista de nomes, códigos-fonte, documentos que são organizados de forma sequencial;
- 2 dimensões - mapas e leiautes de jornais;
- 3 dimensões - objetos do mundo real, como moléculas, pessoas e edifícios;
- Temporais - séries temporais como registros médicos ou de projetos;
- Multidimensionais - bancos de dados relacionais;
- Árvores - dados que tem relações entre si seguindo uma hierarquia;

- Redes - dados que tem relações entre si, mas que não seguem uma hierarquia como nas árvores.

As representações gráficas são as estruturas que expressam a informação contida dentro de um conjunto de dados (JAESCHKE; LEISLER; HEMMJE, 2005). O trabalho de Lohse et al. (1994) elenca diversos tipos de representações gráficas: mapas simbolizando a geografia física de um local, cartogramas representando dados quantitativos espaciais, diagramas estruturados e de processos, redes com suas ligações entre os componentes, tabelas exibindo um conjunto de fatos ou relacionamentos em formato compacto e gráficos demonstrando dados quantitativos.

Com aumento da coleta e uso de dados quantitativos nos últimos séculos, uma grande variedade de gráficos surgiu para suprir a demanda da disposição de informações de diferentes tipos com diferentes intuitos (MACDONALD-ROSS, 1977).

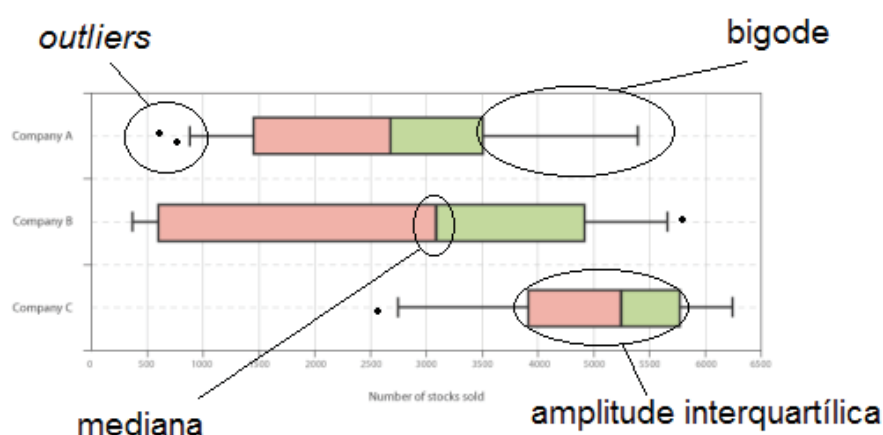
Os gráficos são o tipo de representação mais escolhido para a visualização de informações. De acordo com Collis e Hussey (2003), os gráficos são utilizados para a apresentação de dados de forma precisa. Diversos tipos de gráficos são utilizados:

- *Boxplot*

O *boxplot* (Figura 18), diagrama de John Tukey, tem como função demonstrar estatística descritiva (TUFTS, 2001). Com objetivo de possibilitar a distribuição de um conjunto de dados, o diagrama conta com os chamados bigodes, alinhados verticalmente ou horizontalmente, indicam a variabilidade fora dos quartis.

Os *outliers* são demonstrados através de pontos acima ou abaixo dos bigodes. A caixa entre os bigodes sinaliza os quartis, e o traço que cruza a caixa sinaliza a mediana.

Figura 18 – EXEMPLO DE BOXPLOT

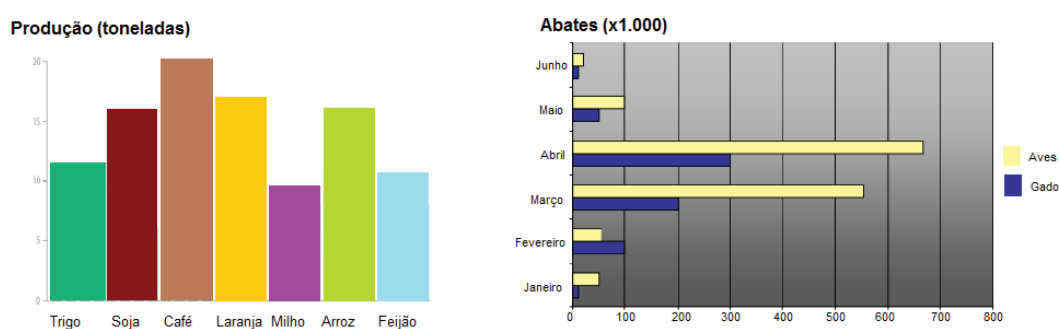


Fonte: o Autor, adaptado de Pereira (2015)

- Gráfico de barras

Os gráficos de barras (Figura 19) contam com variáveis dependentes distribuídas em barras verticais ou horizontais (MACDONALD-ROSS, 1977). Utilizado para comparação de variáveis, o gráfico sempre começa no zero, tendo suas barras espaçadas uma da outra (COLLIS; HUSSEY, 2003).

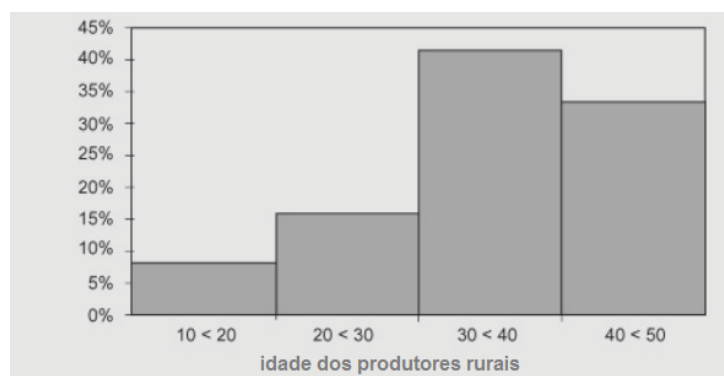
Figura 19 – EXEMPLO DE GRÁFICO DE BARRAS VERTICAL E HORIZONTAL



Fonte: o Autor, adaptado de Pereira (2015) e Dias (2007)

O histograma (Figura 20) caracteriza-se como uma variação do gráfico de barras, onde as barras se tocam com intuito de indicar uma continuidade nos dados (COLLIS; HUSSEY, 2003). Os intervalos das classes devem ser iguais, e ambos os eixos devem conter escalas numéricas (MACDONALD-ROSS, 1977).

Figura 20 – EXEMPLO DE HISTOGRAMA



Fonte: o Autor, adaptado de Collis e Hussey (2003)

- Gráfico de dispersão

Utilizado com o intuito de demonstrar a relação entre duas ou mais variáveis, o gráfico de dispersão (Figura 21) utiliza-se de pontos plotados nos valores das variáveis, sendo útil para análise da correlação entre variáveis (COLLIS; HUSSEY, 2003).

Figura 21 – EXEMPLO DE GRÁFICO DE DISPERSÃO



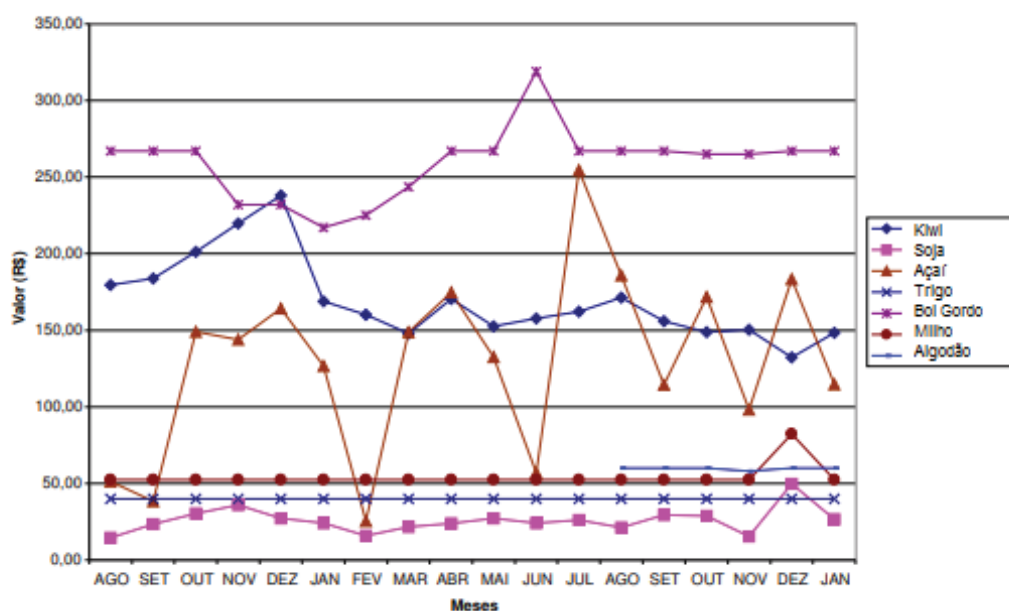
Fonte: Pereira (2015)

- Gráfico de linhas

Este tipo de representação gráfica evidencia oscilações e regularidades, além de demonstrar tendências de redução ou aumento durante um período de tempo, conforme indicado no trabalho de Nascimento e Ferreira (2005).

O gráfico de linhas (Figura 22) trabalha com uma série pontos ligados por uma linha. Este gráfico tem como vantagem, perante aos demais, a possibilidade de sobreposição das linhas no mesmo eixo (COLLIS; HUSSEY, 2003). Os gráficos de linhas possibilitam facilitar a visualização, por parte dos gestores, de informações históricas de indicadores (FRANCISCHINI; FRANCISCHINI, 2017).

Figura 22 – EXEMPLO DE GRÁFICO DE LINHAS

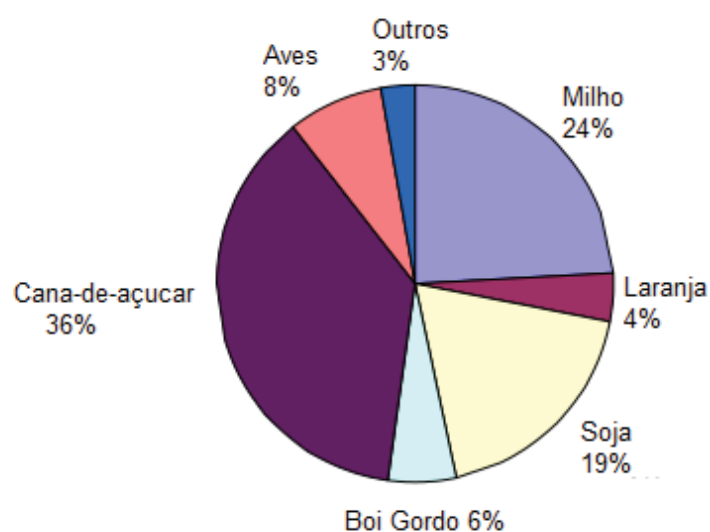


Fonte: o Autor, adaptado de Nascimento e Ferreira (2005)

- Gráfico de pizza

O gráfico de pizza caracteriza-se por ser circular, dividido em segmentos de tamanho proporcional a sua área no diagrama, sendo indicado seu uso para até seis segmentos (COLLIS; HUSSEY, 2003). São utilizados quando se busca demonstrar porcentagens. Seu uso pode ser prejudicado quando porcentagens são pequenas, de cinco por cento para baixo (SILVA NETO, 2008)(Figura 23).

Figura 23 – EXEMPLO DE GRÁFICO DE PIZZA



Fonte: o Autor, adaptado de Pereira (2015)

A combinação de diversos gráficos geram um painel para a disponibilização da informação chamados *dashboards* (DIAS, 2007). As *dashboards* propiciam soluções poderosas para quem necessita de suporte de informações para a tomada de decisão (FEW, 2006).

De acordo com Nutini (2015, p. 105), a demonstração de um sistema de indicadores via *dashboard* simplifica a vida do gestor, pois nela contem tudo que é importante de ser visualizado.

As *dashboards* permitem, simultaneamente, o controle de diversos indicadores através de seus gráficos e diagramas, facilitando a visualização de métricas importantes (FRANCISCHINI; FRANCISCHINI, 2017).

Para melhor funcionamento das *dashboards* (Figura 24), se fazem necessárias interações entre as estruturas, oferecimento de maior número de informações úteis possíveis, e ser intuitiva, criativa e confortável para o usuário.

Figura 24 – EXEMPLO DE DASHBOARD



Fonte: o Autor, adaptado de Few (2006)

As *dashboards* também se caracterizam por trazer as os componentes gráficos em uma única tela, com intuito de facilitar, agilizar a exploração das informações na busca de melhor absorção das mesmas (SHARDA; DELEN; EFRAIM, 2013).

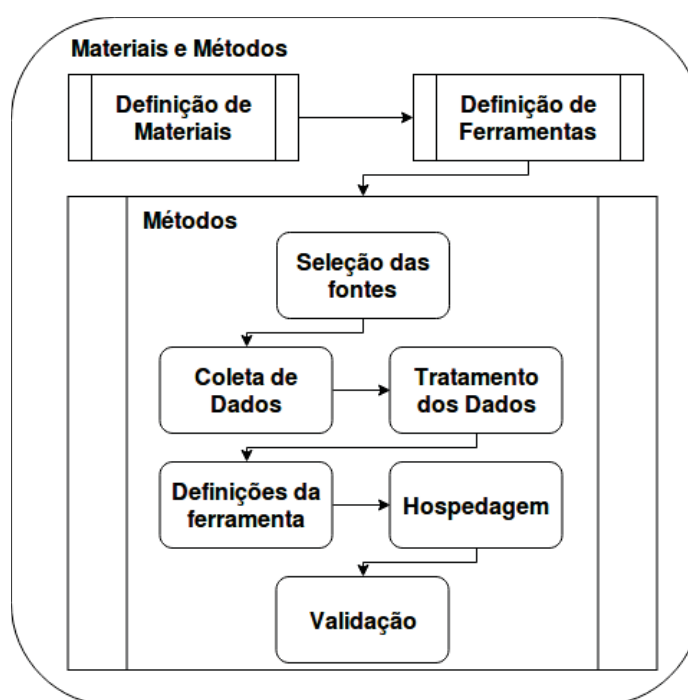
Em seu trabalho, Eckerson (2006, p. 13) também aponta características de uma *dashboard*:

- Propósito - medir de desempenhos;
- Usuários - supervisores, especialistas
- Atualizações - transmitir informações no momento certo
- Dados - eventos
- Exibição - gráficos visuais e dados brutos

3 MATERIAIS E MÉTODOS

A presente seção descreve os materiais e métodos utilizados, sendo dividida em três tópicos principais (Figura 25): descrição de materiais, ferramentas e métodos utilizados para alcance de objetivos elencados na introdução da pesquisa.

Figura 25 – SEQUÊNCIA LÓGICA DOS MATERIAIS E MÉTODOS UTILIZADOS NA DISSERTAÇÃO



Fonte: o Autor (2018)

3.1 MATERIAIS

A construção da ferramenta de visualização de indicadores de produção do agronegócio depende, primeiramente, da obtenção de dados que a fomenta para que possa exercer funções que se propõe a realizar.

A coleta dos dados inicia o ciclo de vida da informação, conforme indicado por Chaffey e Wood (2004). Os dados coletados são vitais para os resultados encontrados pela pesquisa, pois é a partir da aquisição dos mesmos que se pode seguir os passos metodológicos planejados para a construção do trabalho.

Assim como explanado no capítulo Fontes de Dados e Informações (2.4, p. 69), existem diversas classificações de fontes de dados; as que cabem ao presente

trabalho são as externas, de acordo Kobrin et al. (1980), pois vêm de fontes diferentes do autor do trabalho. As fontes buscadas também podem ser classificadas como: de negócios por Montalli (1991) por se tratarem de dados sobre mercados; e de fontes primárias, que, de acordo com Boteon (2004), são organizações públicas e privadas que produzem dados.

O levantamento leva à fontes que disponibilizam dados do agronegócio brasileiro, dados estes classificados como material essencial para o desenvolvimento da ferramenta que o trabalho se propõe a disponibilizar.

3.2 FERRAMENTAS

O desenvolvimento da ferramenta proposta por este trabalho exige a utilização de uma linguagem de programação que possibilite:

- Acesso as fontes disponibilizadoras de dados na internet;
- Manipulação de arquivos e extração automática dos dados disponibilizados pelas fontes;
- Geração do arquivo com dados tabulados de diferentes fontes;
- Interpretação dos dados via gráficos e análise estatística;
- Filtragem dos dados pelo usuário;
- Controle de erros que possam acontecer

Conforme indicado na seção ETL e Linguagens de Programação (2.4.3, p. 65) a linguagem R possibilita realizar as ações necessárias para a construção da ferramenta proposta neste trabalho através de suas bibliotecas de extração, tratamento e disponibilização de dados e controle de erros através da utilização de funções de *try catch*.

Para o desenvolvimento de ferramentas computacionais se faz o necessário o uso de uma Integrated Development Environment (IDE), ambiente de desenvolvimento de algoritmos.

O RStudio é um ambiente de desenvolvimento da linguagem R que de acordo com Verzani (2011), traz como uma de suas vantagens utilização de quatro painéis: console, editor de código-fonte, organizador de documentos e um painel para inclusão de anotações. A disponibilidade em diferentes sistemas operacionais e integração entre editor de código-fonte e console também é apontada como vantagem pelo autor.

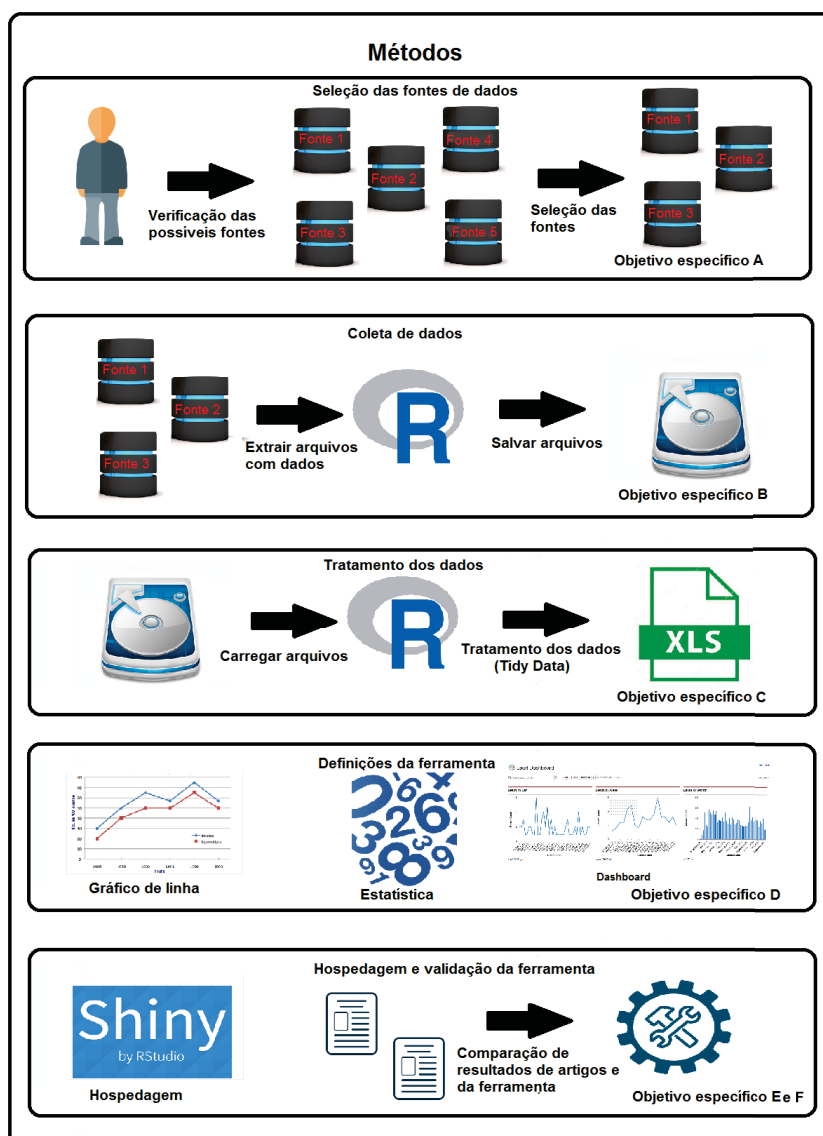
Sendo assim, foi selecionada a linguagem R e suas bibliotecas e a IDE RStudio como ferramentas para a construção dos indicadores do agronegócio do Brasil e cumprimento dos objetivos propostos.

3.3 MÉTODOS

Para a construção do presente trabalho segue-se metodologias indicadas no referencial teórico como: *extract, transform, load* (ETL), utilizada sobretudo para construção de *data warehouses* na extração, transformação e carregamento de informações; processos de transformação da informação indicados por Chaffey e Wood (2004), conceitos de *tidy data* para geração do arquivo com dados transformados e o processo de transformação de dados em representações gráficas de acordo com Card, Mackinlay e Shneiderman (1999).

Os procedimentos metodológicos (Figura 26) foram separados em cinco etapas, que contemplam desde a verificação das possíveis fontes dos dados até a comparação dos resultados da ferramenta e de trabalhos científicos.

Figura 26 – FLUXO DOS MÉTODOS DA PESQUISA



Fonte: o Autor (2018)

3.3.1 Seleção das fontes de dados

O primeiro passo diz respeito à identificação das fontes externas dos dados do agronegócio brasileiro, visitando os endereços eletrônicos das organizações ligadas ao agronegócio (QUADRO 9) indicadas na seção Fontes de Dados do Agronegócio (2.4.5, p. 69), com intuito de identificar a disponibilização de dados que se encaixem no escopo do presente trabalho.

O quadro seguinte foi construído da seguinte forma: foram selecionadas as organizações citadas em mais de um trabalho utilizado durante a construção do Capítulo Referencial Teórico (2, p. 30). Por exemplo, a CONAB encontra-se na listagem de

organizações por ser citada em trabalhos como de Albano, Wohlenberg e Garcia (2014), Baccarin, Gebara e Borges Junior (2011) entre outros, a EMBRAPA, por ser mencionada por Abbade (2014), Albano, Wohlenberg e Garcia (2014) e etc, e assim sucessivamente.

QUADRO 9 – ENDEREÇO ELETRÔNICO DE ORGANIZAÇÕES LIGADAS AO AGRO-NEGÓCIO

Organização	Endereço eletrônico consultado
Companhia Nacional de Abastecimento (CONAB)	https://www.conab.gov.br/
Confederação da Agricultura e Pecuária do Brasil (CNA)	http://www.cnabrazil.org.br/
Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA)	https://www.embrapa.br/
Instituto Brasileiro de Geografia e Estatística (IBGE)	https://www.ibge.gov.br/
Ministério da Agricultura, Pecuária e do Abastecimento (MAPA)	http://www.agricultura.gov.br/
Ministério da Indústria, Comércio Exterior e Serviços	http://www.mdic.gov.br/
Organização das Nações Unidas para Agricultura e Alimentação (FAO)	http://www.fao.org/home/en

Fonte: o Autor (2018)

Os dados inseridos no modelo de extração, devem atender: a delimitação da pesquisa, que indica dados quantitativos não-financeiros da fase "dentro da porteira", a possibilidade de extração via *script* R, linguagem selecionada para construção da ferramenta, e apontados na literatura para validação teórica do dado.

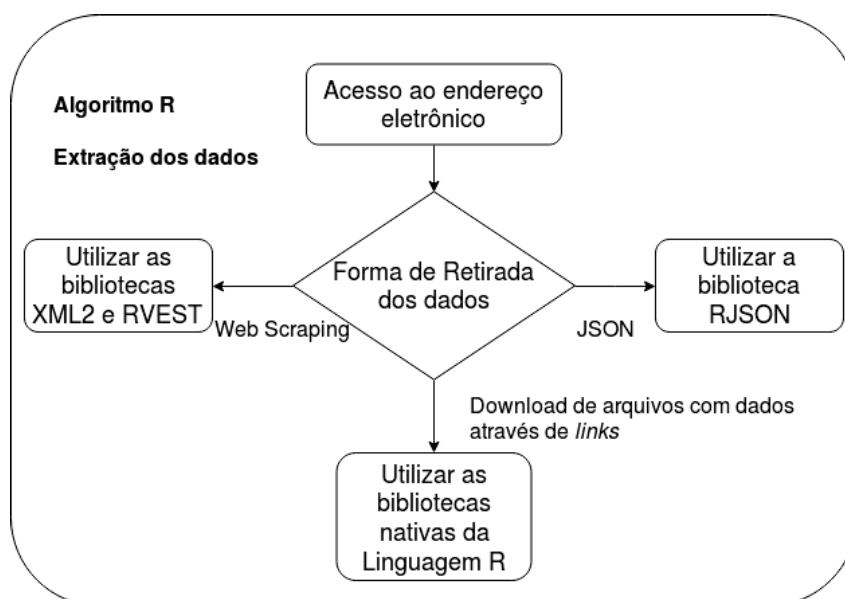
3.3.2 Coleta de Dados

O processo de transformação da informação, caracterizado por Chaffey e Wood (2004), inicia-se através da coleta de dados sendo sucedida pela organização e processamento posteriores. Os procedimentos aqui descritos se caracterizam como extração de dados, o primeiro passo da metodologia ETL.

A coleta dos dados depende de técnicas selecionadas e do instrumento de coleta (MARCONI; LAKATOS, 2007). A linguagem R, através de seu conjunto de instruções indicados anteriormente (seção 2.4.3 - ETL e Linguagens de Programação, p. 65), possibilita a coleta de dados de fontes documentais externas, fazendo com que o algoritmo em R gerado exclusivamente para atender o caso desta pesquisa seja o instrumento da coleta.

A indicação do algoritmo instrumento de coleta (Figura 27) construído, com o objetivo de realizar a extração automática dos dados faz com que seja cumprido o objetivo específico da geração do modelo de extração dos dados.

Figura 27 – FLUXOGRAMA MODELO DO ALGORITMO DE EXTRAÇÃO DOS DADOS



Fonte: o Autor (2018)

O algoritmo trabalha com diferentes técnicas para a coleta dos dados:

- *Web Scraping* - esta técnica possibilita que os dados sejam coletados diretamente da página da organização, sem que seja necessário baixar nenhum arquivo;
- JSON - permite que os dados sejam obtidos diretamente de uma página *web* no formato JSON para que sejam tratados posteriormente;
- Download de arquivos - o algoritmo em R acessa o endereço eletrônico da organização e realiza a requisição para *download* do arquivo com os dados desejados.

3.3.3 Tratamento dos dados

O tratamento dos dados é realizado seguindo conceitos de *tidy data*, por ser recomendado para se utilizar em ferramentas de análise de dados.

No que diz respeito ao tratamento dos dados, o *script* R desenvolvido para este trabalho (Apêndice E) e responsável pelo tratamento dos dados de todas as fontes encontradas durante a coleta, realiza:

- Filtragem dos dados - retira dados com valores nulos ou zerados, para tornar o arquivo com os dados mais leve;

- Transformação dos dados - formatação dos dados para que fiquem com a mesma unidade de medida, nome de indicador e etc;
- Agregação dos dados - junta registros de determinados estados para gerar dados de uma região. Por exemplo: juntar a produção de milho dos estados do Paraná, Santa Catarina e Rio Grande do Sul do ano de 2015 para gerar a produção total de milho da região Sul em 2015. Reúne todos os dados de diferentes fontes em um único arquivo.

A realização da filtragem, transformação e agregação servem para que os dados, oriundos de diferentes fontes, possam ser armazenados em um único arquivo, utilizado posteriormente para visualização dos dados. Os dados são inseridos em uma tabela com a estrutura apresentada no QUADRO 10.

QUADRO 10 – COLUNAS DA PLANILHA DE DADOS DO AGRONEGÓCIO

Coluna	Descrição	Exemplo
Ano	Indica o ano da ocorrência. Variável quantitativa discreta	1980, 1996, 2015
Local	Indica o local da ocorrência. Pode representar a ocorrência de um estado ou do Brasil. Variável nominal	Paraná, São Paulo, Ceará, Brasil
Fonte	Fonte da informação de quais os dados foram extraídos. Variável nominal	IBGE, CONAB
Produto	Nome do produto que foi medido o indicador. Variável nominal	Abacaxi, café, milho
Indicador	Nome do indicador medido. Variável nominal	Área plantada, área colhida, produtividade
Unidade de Medida	Indica qual a unidade de medida utilizada na medição do indicador. Variável nominal	Mil toneladas, mil hectares mil litros
Valor	Valor da observação. Variável quantitativa contínua	12.5, 0.7, 156,8

Fonte: o Autor (2018)

A possibilidade de geração de uma planilha formato .XLSX que siga conceitos de *tidy data* pode ser interpretada pela biblioteca *Shiny*. Isso fez com que, neste caso, se optasse pela não utilização de banco de dados, pois seu uso envolveria a modelagem de um banco de dados e construção de uma série de instruções para inserção, seleção e edição dos dados, o que tornaria a ferramenta mais complexa no que diz respeito a sua codificação e não traria vantagem no uso ou processamento.

Após a realização das ações de filtragem, transformação e agregação dos dados, é gerado um único arquivo com todos os dados (linhas 1.158 e 1.159 do Apêndice E), e assim se cumpre o objetivo específico de criação do modelo de padronização dos dados das diferentes bases.

Para a constatação de possíveis erros, tanto na parte de coleta quanto no tratamento dos dados, o algoritmo se utiliza de *try catches* para que um arquivo de log de erros seja gerado caso algum erro ocorra.

3.3.4 Definições da ferramenta

Seguindo o modelo de transformação da informação indicado por Chaffey e Wood (2004), os dados coletados deverão ser transformados em informação através do que os autores chamam de processos de transformação.

Neste trabalho, o processo de transformação selecionado foi a Visualização da Informação (VI), presente na seção 2.5 (p. 70), através da ciclo indicado por Card, Mackinlay e Shneiderman (1999), que indica a realização de um mapeamento visual de uma tabela de dados para se gerar estruturas visuais.

A VI é capaz de fazer com que o usuário perceba a informação contida nos dados com maior naturalidade e velocidade, além de possibilitar a interação do usuário com a informação disposta conforme levantado na seção 2.5, p. 70.

A ferramenta utilizará funções da biblioteca *Shiny* da linguagem R para demonstração de estruturas visuais e de filtros para interação do usuário. Para a geração das estruturas visuais, selecionou-se as seguintes características para serem utilizadas na ferramenta:

- Gráficos de linha

Conforme levantado na seção Visualização da Informação (2.5), a possibilidade de sobreposição das linhas, trabalho com séries históricas e com indicadores fez com que se definisse a utilização deste tipo de gráfico.

Tomou-se também como base a utilização do gráfico de linha a partir de trabalhos como de Gasques et al. (2004), Francischini e Francischini (2017), Caldeira (2016) e Contini et al. (2006), que são exemplos de utilização de gráficos deste tipo em estudos ligados à indicadores e ao agronegócio.

Para cada valor diferente encontrado na coluna "indicador" é gerado um gráfico de linha com seus respectivos dados. Os gráficos são construídos pela biblioteca Plotly, através de sua função "ggplotly", que constrói gráficos com opções nativas como *zoom* e a possibilidade de esconder linhas do gráfico.

- Métodos estatísticos - Estatística Descritiva e *Boxplot*

O fato de trabalhar com dados quantitativos abre-se o precedente da utilização de métodos estatísticos. A estatística descritiva visa descrever o conjunto de dados que se analisa, facilitando o entendimento dos dados que estão diante do usuário. Fazem parte dos métodos estatísticos descritivos presentes neste trabalho as medidas de: média, desvio padrão, mínimo e máximo. As funções nativas da linguagem R `mean`, `sd`, `min` e `max` se encarregam de realizar os cálculos.

O *boxplot* será utilizado pelo fato de auxiliar a demonstração de medidas estatísticas, tornando-se uma medida visual de medidas de dispersão.

Os *boxplots* serão apresentados logo após os gráficos de linhas, agrupando dados de cada uma das décadas, por exemplo, um *boxplot* contemplando dados de 1970 até 1979, ao lado outro *boxplot* de 1980 até 1989 e assim sucessivamente.

- Filtros

A geração da tabela com os dados tratados possibilitará elencar filtros para implementação na ferramenta, sendo que as colunas ano, local, fonte e produto, mapeadas no QUADRO 10 (p. 87), entrarão no rol de filtros disponíveis dentro da ferramenta para livre interação do usuário.

- *Dashboard - Tableau de Bord*

A utilização do conceito de *dashboard*, assim como utilizado na ferramenta da Companhia Nacional do Abastecimento (CONAB), possibilita fornecer ao usuário o controle diversas métricas de uma vez através de vários gráficos em uma mesma tela, o que melhora a experiência do usuário.

Assim como *dashboard*, o sistema de medição de desempenho Tableau de Bord permite a visualização de diversos indicadores ao mesmo tempo em formas gráficas, alinhando-se com o objetivo deste trabalho.

Quanto aos metadados da ferramenta, selecionou-se o esquema Dublin Core para representação dos mesmos, utilizando-se dos seguintes elementos: título, assunto, relações, fonte, cobertura, tipo, descrição, criador, colaborador, data, formato e idioma.

Alguns elementos do rol original do Dublin Core não serão utilizados: colaborador pelo fato do criador e editor serem a mesma pessoa; direitos pela ferramenta não ser patenteada; e identificador por não haver nenhuma referência à ferramenta em algum contexto. Os demais elementos seguem o padrão indicado no QUADRO 11.

QUADRO 11 – METADADOS DA FERRAMENTA

Elemento	Descrição	Valor
Título	Nome da ferramenta	Indicadores de Produção do Agronegócio Brasileiro
Assunto	Palavras-chave ligadas a ferramenta	Agronegócio, Indicadores de Produção, Visualização da Informação
Relações	Relações da ferramenta proposta com ferramentas existentes	CONAB - Portal de Informações Agropecuárias Observatório Agrícola EMBRAPA - AGROPENSA
Fonte	Fontes dos dados	CONAB, EMBRAPA (exemplo)
Cobertura	Cobertura espacial da ferramenta; padrão <i>The Getty Thesaurus of Geographic Names</i> (TGN) de georreferenciamento	1000047
Tipo	Tipo dos dados	Dados relacionados a produção do agronegócio brasileiro
Descrição	Descrição textual da ferramenta	A ferramenta tem como objetivo disponibilizar dados do agronegócio brasileiro através de gráficos e filtros interativos
Criador	Nome do autor	Jhonny Ivair de Lima Maciel
Colaborador	Nome do orientador	Prof. Dr. Egon Walter Wildauer
Data	Data da última atualização dos dados; ISO 8601	2018-10-10 (exemplo)
Formato	Formato do recurso	Ferramenta web em R utilizando a biblioteca <i>Shiny</i>
Idioma	Idioma o qual a ferramenta utiliza; ISO 639	pt

Fonte: o Autor (2018)

Alguns metadados como data e fonte não podem ser definidos previamente; a data, é um dado que será alterado conforme atualizações dos dados da ferramenta e a declaração das fontes, depende do cumprimento do primeiro objetivo específico da presente pesquisa, que diz respeito a identificação das fontes dos dados.

Tarefas como a coleta e limpeza de dados e publicação de uma nova versão da ferramenta são realizadas diariamente, de modo que os dados sempre estão atualizados para sua utilização. O Cron, *software* de sistemas operacionais UNIX e Linux, possibilita o agendamento de tarefas através do Crontab, que é responsável por realizar a edição do arquivo em que são especificados quais as ações a serem executadas pelo sistema operacional e quando isso ocorre (PETERS, 2009).

No caso do presente trabalho, a ação executada pelo sistema operacional é o *script* que realiza a coleta e a limpeza dos dados relacionados ao agronegócio e a

publicação da ferramenta no ShinyApps. A rotina inicia-se sempre as 00:00 de cada dia, todos os dias da semana, para que os dados estejam sempre o mais atualizados possível.

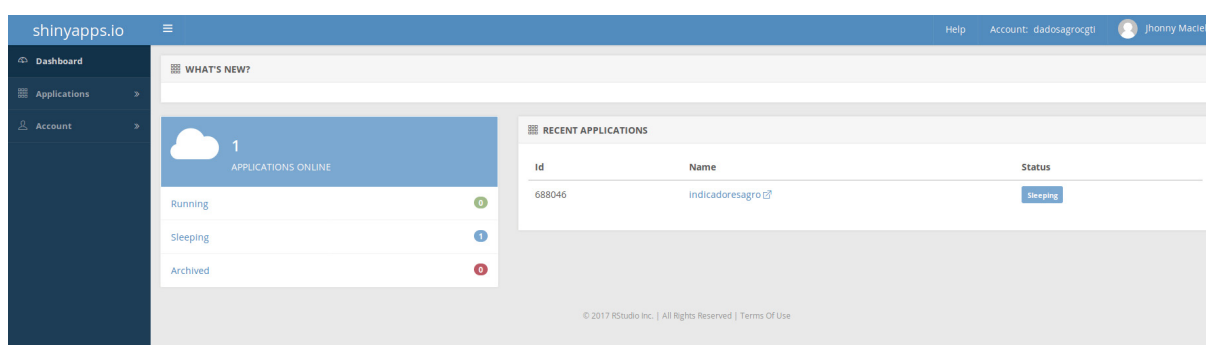
3.3.5 Hospedagem da ferramenta

Conforme indicado na seção sobre ferramentas utilizadas, a biblioteca *Shiny* disponibiliza o suporte para hospedagem de ferramentas de maneira gratuita. Optou-se por hospedar a ferramenta dentro do serviço de hospedagem de aplicações Shiny em nuvem ShinyApps¹ servidor da biblioteca por ser uma própria indicação da Shiny.

Após um rápido cadastro na página de acesso ao ShinyApps², é possível criar uma conta, e através do nome de conta definido, acessar as ferramentas hospedadas para esta conta. Por exemplo, o nome da conta definida para o presente trabalho chama-se "dadosagrocgti", e o endereço para acesso à ferramenta é o <https://dadosagrocgti.shinyapps.io/indicadoresAgro>. O "indicadoresAgro" diz respeito ao nome da ferramenta hospedada no ShinyApps.

Após a criação da conta e o acesso ao ShinyApps, o usuário encontra uma *dashboard* onde a primeira vista são apresentadas as ferramentas hospedadas para a conta logada (Figura 28). Ao clicar no nome de uma das ferramentas, o usuário pode realizar ações como resetar, arquivar, restaurar e remover uma aplicação, visualizar estatísticas do uso da ferramenta e alterar algumas configurações.

Figura 28 – DASHBOARD SHINYAPPS



Fonte: o Autor (2018)

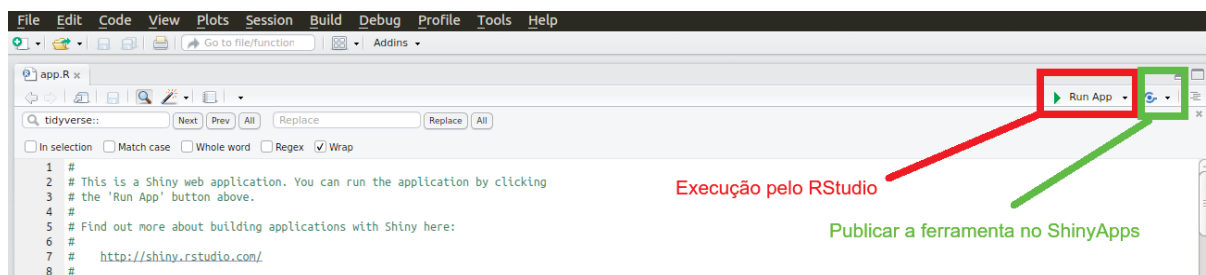
Através da instalação de um complemento dentro do RStudio, é possível realizar a publicação da ferramenta no ShinyApps dentro da própria IDE. A instalação se dá ao criar um app do Shiny, ao verificar que ainda não há a instalação, o RStudio solicita sua instalação.

¹ <<https://www.shinyapps.io/>>

² <<https://www.shinyapps.io/admin/#/login>>

Após a instalação, o RStudio permite que aplicações em Shiny sejam executadas pela própria IDE ou também que sejam publicadas diretamente no ShinyApps (Figura 29).

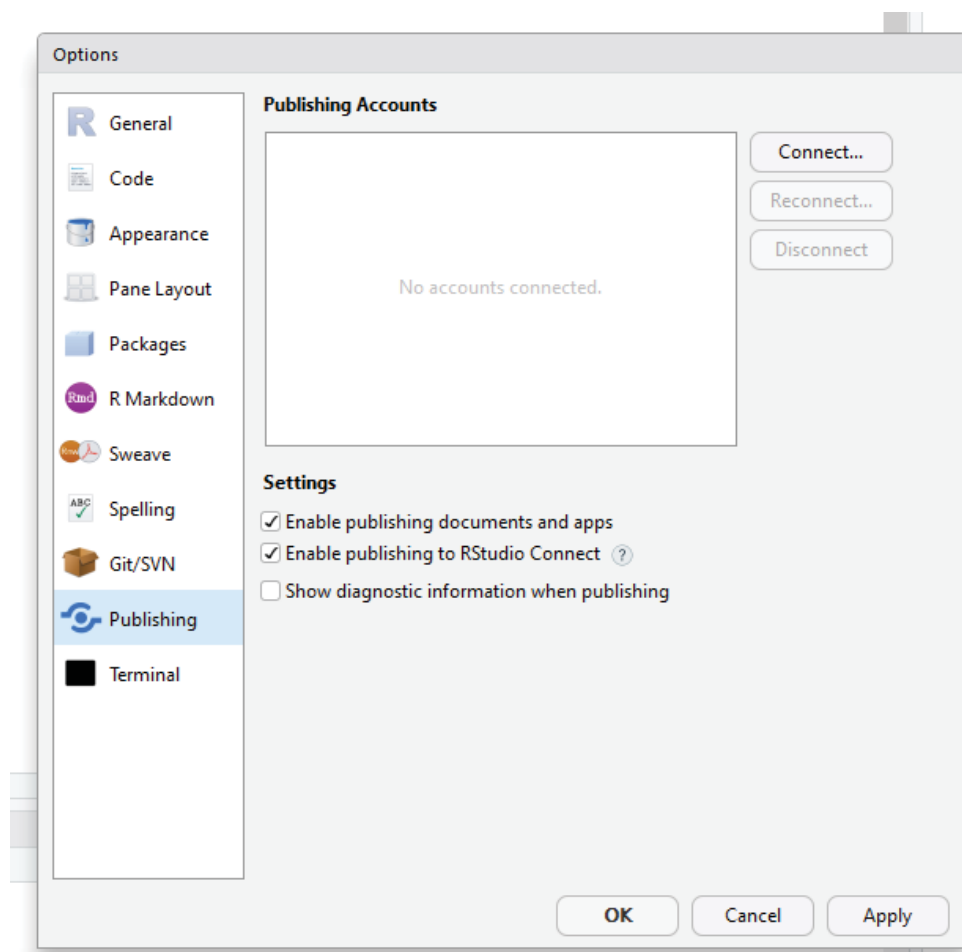
Figura 29 – EXECUÇÃO DE APLICATIVOS SHINY NO RSTUDIO



Fonte: o Autor (2018)

Antes de realizar a publicação da aplicação é necessário realizar a sincronização de uma conta ShinyApps com o RStudio. Para isso, dentro do ícone do ShinyApps existe um menu chamado "Manage Account", onde é realizada a sincronização (Figura 30).

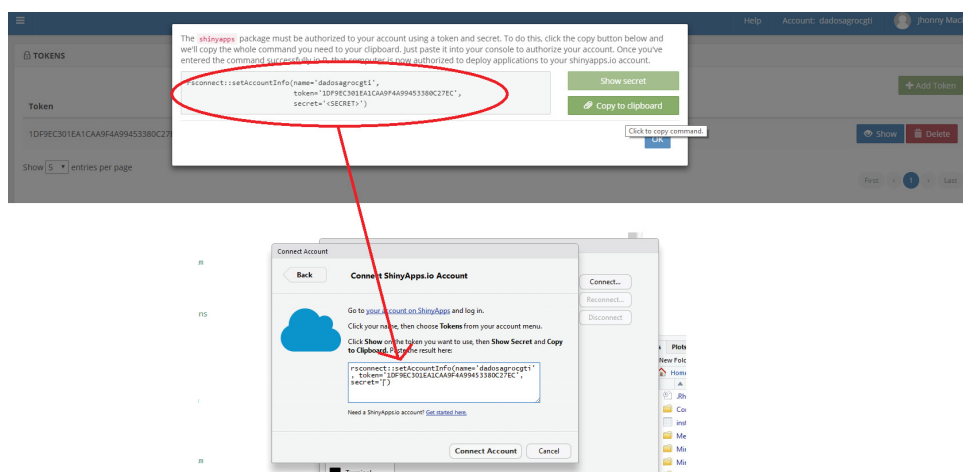
Figura 30 – CONTROLE DE CONTAS SHINYAPPS LIGADAS AO RSTUDIO



Fonte: o Autor (2018)

Ao selecionar o botão "connect" e a sincronização com o ShinyApps, o RStudio solicita um *token*, disponibilizado dentro de cada conta criada no ShinyApps, localizada no menu "Account" e depois no submenu "Token". Dentro desta página é possível, clicando no botão "Show", copiar o *token* necessário para vinculação da conta ShinyApps com o RStudio (Figura 31).

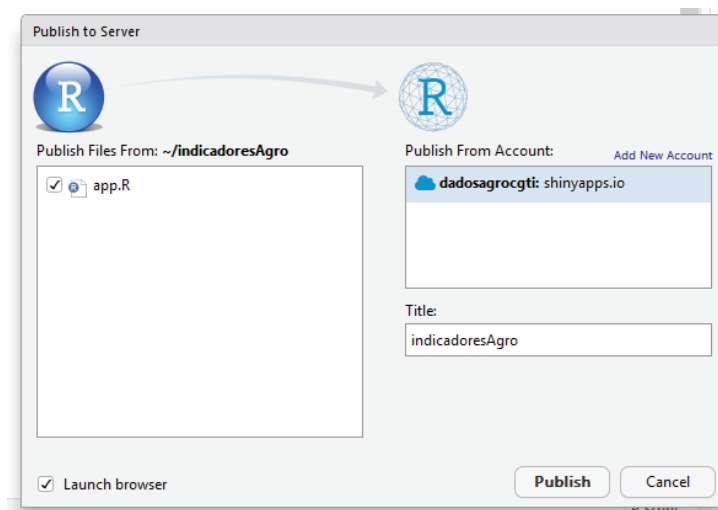
Figura 31 – INSERÇÃO DO TOKEN DO SHINYAPPS NO RSTUDIO



Fonte: o Autor (2018)

Após a vinculação, torna-se possível publicar os aplicativos automaticamente via RStudio (Figura 32).

Figura 32 – PUBLICAÇÃO DE APLICAÇÃO NO SHINYAPPS VIA RSTUDIO



Fonte: o Autor (2018)

Além da publicação pela interface do RStudio, é possível também realizar esta ação através de instruções dentro de um algoritmo em R. Para isso são necessários dois passos: conexão com a conta do ShinyApps e a publicação da ferramenta.

Para a conexão, é necessário utilizar a função `rsconnect::setAccountInfo`, indicando como parâmetro o nome do usuário do ShinyApps, o *token* e o *secret* disponibilizados pela ferramenta, como apresentado na parte superior da Figura 31

(linhas 43 à 45 do Apêndice E). Para realizar a publicação, basta utilizar a função "deployApp", passando como parâmetros: o diretório onde se encontra o que deve ser publicado, os nomes dos arquivos e o nome da ferramenta a ser hospedada (linha 1.561 do Apêndice E).

3.3.6 Comparação de resultados da ferramenta e da literatura

A validação dos dados da ferramenta consiste na comparação de estudos indicados neste trabalho sobre a produção de quaisquer produtos do agronegócio com os resultados indicados pela ferramenta proposta no trabalho, com intuito de revelar se os resultados indicados apresentam variações e se são convergentes.

Dos trabalhos utilizados na construção do Referencial Teórico do presente trabalho, sete apresentam indicadores de desempenho de produção e serviram para a comparação dos da ferramenta e da literatura (QUADRO 12).

QUADRO 12 – TRABALHOS UTILIZADOS NA VALIDAÇÃO DOS DADOS FERRAMENTA

Título	Autor
Caracterização da produção de cana-de-açúcar através de indicadores espaciais e temporais em Mato Grosso do Sul	Maia et al. (2015)
Expansão Canavieira e Ocupação Formal em Empresas Sucroalcooleiras do Centro-Sul do Brasil, entre 2007 e 2009	Baccarin, Gebara e Borges Junior (2011)
Evolução recente e tendências do agronegócio	Contini et al. (2006)
Análise da competitividade do algodão mato-grossense no cenário internacional: 1999 a 2005	Sousa, Bonjour e Figueiredo (2006)
Agronegócio, produção de alimentos e segurança alimentar na América Latina	Costa (2016)
O agronegócio no Brasil e a produção capitalista do território	Ferreira (2012)
Desempenho e crescimento do agronegócio no Brasil	Gasques et al. (2004)

Fonte: o Autor (2018)

4 RESULTADOS

Este capítulo apresenta os resultados obtidos após a realização dos passos metodológicos apontados no capítulo anterior, iniciando pela verificação das possíveis fontes de dados do agronegócio brasileiro, passando pela construção do algoritmo de extração e tratamento dos dados e finalizando com a validação da ferramenta gerada.

4.1 SELEÇÃO DAS FONTES DE DADOS

Para o levantamento das fontes de dados, indicado no capítulo Materiais e Métodos, seção intitulada Seleção das fontes de dados (3.3.1, p. 84), foram verificados os portais eletrônicos de sete organizações como possíveis provedores de dados (QUADRO 9, p. 85). A visita dos portais eletrônicos para discriminação de dados que foram encontrados ocorreu entre os dias 16 e 30 de julho de 2018.

4.1.1 Companhia Nacional do Abastecimento

No endereço eletrônico da Companhia Nacional de Abastecimento (CONAB) é possível encontrar diversos tipos de indicadores ligados ao agronegócio brasileiro, sempre acompanhados de ferramentas gráficas que possibilitam a interação do usuário. Um dos tipos de dados fornecidos pela CONAB diz respeito a produção agrícola, sendo dados quantitativos não-financeiros, atendendo o que foi delimitado no capítulo de Materiais e Métodos.

A CONAB disponibiliza 33 planilhas com os dados de produtos do agronegócio. Dentre estas, algumas delas são do mesmo produto, divididos pelas duas safras anuais do produto e uma que contém o somatório geral da produção anual das duas ou três safras ou sendo divididas pelo tipo do produto (Figura 33).

Figura 33 – DADOS DISPONIBILIZADOS PELA CONAB

<p>02/05/18 17h17</p> <p>Amendoim 1ª safra</p> <p>Amendoim1aSerieHist.xls (52 Downloads)</p>	<p>02/05/18 17h15</p> <p>Arroz Irrigado</p> <p>ArrozIrigadoSerieHist.xls (67 Downloads)</p>
<p>02/05/18 17h16</p> <p>Amendoim 2ª safra</p> <p>Amendoim2aSerieHist.xls (48 Downloads)</p>	<p>02/05/18 17h14</p> <p>Arroz Sequeiro</p> <p>ArrozSequeiroSerieHist.xls (64 Downloads)</p>

Fonte: CONAB (2018)¹

Das 33 planilhas disponibilizadas pela CONAB, foram selecionadas 25 para serem utilizadas neste projeto (Apêndice A). Dentre as que foram descartadas, duas são relacionadas a produção nacional, duas a produção de cana-de-açúcar e quatro relacionadas a produção de café. Quanto as nacionais, uma delas indica o total de produção do Brasil produto por produto, e a outra estado por estado. Por estes dados já estarem contidos em outras planilhas, não serão utilizados.

Quanto aos dados de cana-de-açúcar não utilizados, a planilha chamada pela CONAB de "Cana-de-Açúcar - Indústria" apresenta dados relativos a indústria, e não a fase "dentro da porteira", que é o escopo do projeto. Já a planilha "Cana-de-Açúcar - Área Total" aponta dados não levantados na literatura, como por exemplo: área de mudas, área de expansão, área de renovação e etc. A planilha "Café Total - Parque Cafeeiro" também contempla indicadores que não se enquadram na presente pesquisa: pés em formação, pés em produção e pés total. Os arquivos com dados de Café Arábica, Conilon e total também informam indicadores fora do escopo: área em produção, área em formação, produtividade em sacas por hectare e produção em sacas beneficiadas, que acaba configurando a fase "depois da porteira".

A CONAB disponibiliza dados de três tipos de indicadores: produção, produtividade agrícola e de área plantada para todos os produtos. Os primeiros dados datam de 1976 e vão até uma previsão do ano atual, contemplando sempre os resultados dos 27 estados brasileiros, do Brasil e por região. Os dados são disponibilizados em arquivos .XLS.

Além dos dados sobre a produção agrícola que servem a este trabalho, a CONAB também disponibiliza dados acerca de estoques públicos, armazenagem, importação e exportação, custos de produção e preços agropecuários, mínimos e de frete em seu endereço eletrônico.

4.1.2 Confederação da Agricultura e Pecuária do Brasil

A Confederação da Agricultura e Pecuária do Brasil (CNA) disponibiliza, no portal eletrônico consultado, apenas o indicador financeiro de preço de *commodities* como café, trigo, soja e milho. Os dados são divulgados diariamente pela plataforma, que possibilita a seleção da *commoditie* e da região em que se deseja visualizar o preço diário. Por se tratarem de dados financeiros, os mesmos fornecidos pela CNA não serão utilizados na presente pesquisa, por não estarem incluídos no escopo.

4.1.3 Empresa Brasileira de Pesquisa Agropecuária

A Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), assim como a CONAB, contempla em seu portal eletrônico uma ferramenta de visualização da

informação chamada Agropensa², que trabalha com dados relacionados a agricultura e pecuária brasileira. Os dados são separados em quatro tópicos principais: produção agrícola, pecuária, abate e comércio exterior.

Dentro dos dados da produção agrícola, a EMBRAPA divulga dados de área colhida, produção e produtividade de 79 produtos de 1990 até 2016.

Dentre os dados de pecuária, constam dados de efetivo de rebanhos, produtos de origem animal, aquicultura, vacas ordenhadas e ovinos tosquiados. Os dados relacionados aos rebanhos se dividem em oito tipos diferentes de animais, e datam de 1974 até 2016.

A segunda parte dos dados relacionados à pecuária dizem respeito a produtos de origem animal, totalizando seis produtos. São disponibilizados dados de produção, valor da produção e preço médio de origem do animal. Por não se encaixarem no escopo do trabalho, os dados de valor da produção e preço médio de origem do animal não serão utilizados.

Dados sobre a aquicultura contemplam a terceira parte dos dados sobre a pecuária. São disponibilizados dados de produção, valor da produção e o preço médio da aquicultura de 24 produtos, de 2013 a 2016. Somente os dados de produção serão utilizados neste trabalho, por se encaixarem dentro do escopo da pesquisa.

A produção de leite contempla o penúltimo grupo de dados informados pela EMBRAPA na seção de pecuária do sistema Agropensa. Dados sobre a quantidade de leite produzida, número de vacas ordenhadas e a média de produção por animal são contabilizados de 1974 até 2016.

A seção de dados de ovinos tosquiados, a última da seção de pecuária, também contempla o número total de animais, quantidade de lã produzida e a média obtida de cada animal. Os dados também datam de 1974 até 2016.

Os dados relacionados ao abate disponibilizados pela EMBRAPA são subdivididos em seis categorias: abate, couro adquirido para curtimento, couro adquirido por terceiros, couro curtido, leite e ovos de galinha.

Com relação ao abate, são disponibilizados dados da quantidade total de abate e o peso total obtido das carcaças de sete diferentes tipos de animais entre 1997 a 2017.

Os dados relacionados à produção de couro (couro adquirido para curtimento, couro adquirido por terceiros e couro curtido) dizem respeito a dados relacionados a fase "depois da porteira", pois os fornecedores de couro utilizados nos dados de couro para curtimento e de terceiros são: intermediários, matadouros frigorífico e municipal,

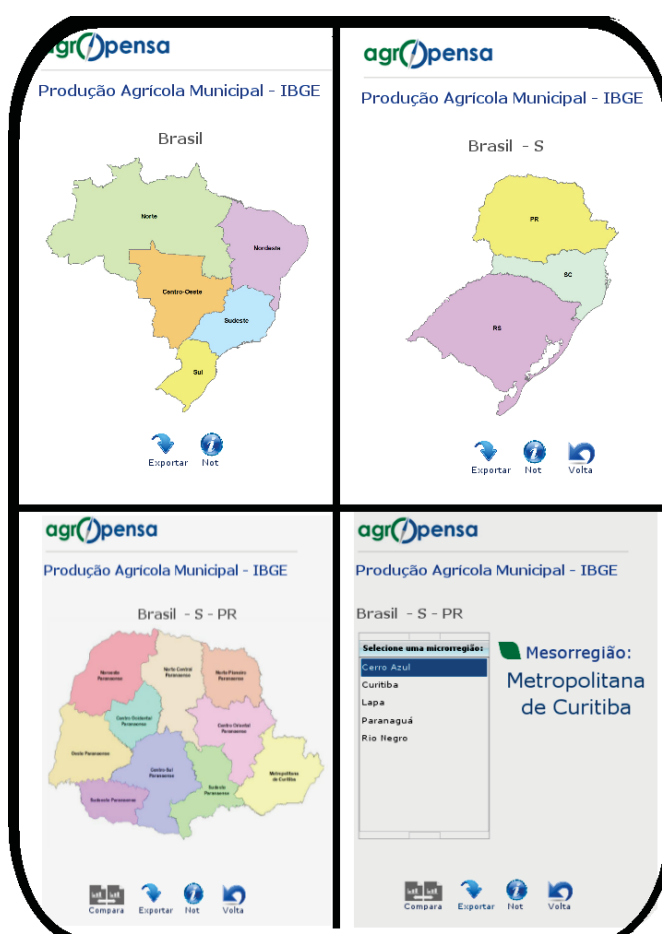
² <<https://www.embrapa.br/agropensa>>

outras origens, outros cortumes e terceiros. Os dados relacionados ao couro curtido dizem respeito a indústria de processamento, conforme indicado na seção da Cadeia Produtiva do Agronegócio no Brasil (2.1.2, p. 37).

Assim como na seção de pecuária, dados de produção de leite também são divulgados na seção de pecuária. A quantidade de leite obtida e industrializada de 1997 até 2017 é disponibilizada pela EMBRAPA. Os últimos dados disponibilizados na seção de abate diz respeito ao número de galinhas poedeiras e de ovos produzidos.

A última seção de dados do Agropensa, da EMBRAPA, diz respeito ao Comércio Exterior. Os dados indicados são de cunho financeiro, não se enquadrando dentro do escopo da presente pesquisa. A EMBRAPA permite a visualização de todos seus dados tanto a nível nacional, regional, estadual e até municipal. (Figura 34).

Figura 34 – SELEÇÃO DE LOCAL PARA EXPORTAÇÃO DE DADOS DA EMBRAPA



Fonte: EMBRAPA (2018)³

A figura mostra os quatro níveis em que o Agropensa permite exportar os dados: nacional, regional, estadual e municipal; da esquerda para direita.

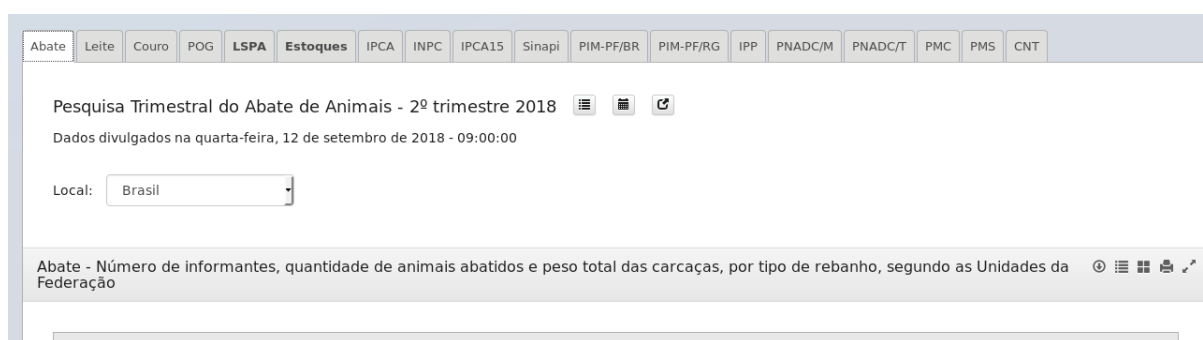
A listagem de todos os produtos que terão seus dados coletados da EMBRAPA encontra-se no Apêndice B.

4.1.4 Instituto Brasileiro de Geografia e Estatística

Através do sistema chamado SIDRA (Sistema IBGE de Recuperação Automática), o Instituto Brasileiro de Geografia e Estatística (IBGE) também disponibiliza dados relacionados a produção do agronegócio brasileiro. Tanto dados da agricultura quanto da pecuária são divulgados pela plataforma.

O IBGE disponibiliza seus dados relacionados ao agronegócio dividindo-os em diversas abas, dentre elas, dados de abate de animais, relacionados a produção de leite, couro, ovos de galinha e de safra (Figura 35).

Figura 35 – DISPONIBILIZAÇÃO DE DADOS NO PORTAL SIDRA - IBGE



Fonte: IBGE (2018)

Além da disponibilização de dados já pronta em cada uma das abas, a ferramenta do IBGE permite que o usuário monte suas próprias consultas em sua base para buscar registros específicos de acordo com a necessidade.

Quanto aos dados relacionados a agricultura, a ferramenta permite buscar dados relacionados a produção, área plantada, área colhida e produtividade de 44 diferentes produtos de 2006 até 2018, e de todos os estados brasileiros.

Nos dados relacionados a pecuária, o IBGE disponibiliza informações sobre abate, leite, couro e produção de ovos. Na seção de abate encontram-se: total de cabeças abatidas, peso total das carcaças e número de informantes dos dados. Por serem dados relacionados a produção e estarem apontados pela literatura utilizada neste trabalho, somente dados de quantidade de cabeças abatidas e o peso total das carcaças serão utilizados. Os dados datam de 1997 até 2018, e são referentes a sete diferentes tipos de animais.

Na seção relacionada à produção de leite são disponibilizados dados dos totais produzidos de 1997 até 2018. Quanto a produção de ovos, a seção consta com dados do total de galinhas poedeiras e o total de ovos produzidos de 1987 a 2018.

Quanto aos dados relacionados ao couro, estes se caracterizam iguais aos dados de couro disponibilizados pela EMBRAPA, caracterizando-se como da fase "depois da porteira", e por se tratarem de dados de um escopo diferente da fase "dentro da porteira" não serão utilizados.

Assim como a CONAB, o IBGE também disponibiliza uma série de outros dados, como índices de preços e uma série de outros indicadores financeiros que não se encaixam no escopo da presente pesquisa.

A listagem dos produtos que terão seus dados coletados pela ferramenta encontra-se no Apêndice C.

4.1.5 Ministério da Agricultura, Pecuária e Abastecimento

O Ministério da Agricultura, Pecuária e Abastecimento (MAPA) disponibiliza em seu endereço eletrônico uma série de *links*⁴ para sistemas da organização. Porém, para a visualização das informações contidas dentro ferramentas indicadas é necessário ser credenciado, o que impossibilita a coleta de dados que poderia ser encontrada a partir desta fonte.

4.1.6 Ministério da Indústria, Comércio Exterior e Serviços

Dentro da seção da balança comercial brasileira⁵ no portal eletrônico do Ministério da Indústria, Comércio Exterior e Serviços há a disponibilização de dados de produtos, inclusive do agronegócio de 1998 até 2018. Por se tratarem de dados financeiros que não se enquadram dentro da delimitação da pesquisa, os dados divulgados não serão utilizados.

4.1.7 Organização das Nações Unidas para Agricultura e Alimentação

A Organização das Nações Unidas para Agricultura e Alimentação (FAO) em seu endereço eletrônico⁶ possibilita ao usuário combinar filtros para geração um arquivo com dados oriundos da combinação em um arquivo .CSV, similar ao sistema do IBGE. Os filtros são de países, produtos, período temporal e as medidas desejáveis.

A seção de dados de produção conta com levantamentos de: cultivo, cultivo processado, animais vivos, produtos primários e processados de origem animal, índices

⁴ <<http://sistemasweb.agricultura.gov.br/>>

⁵ <<http://www.mdic.gov.br/comercio-exterior/estatisticas-de-comercio-exterior/balanca-comercial-brasileira-acumulado-do-ano>>

⁶ <<http://www.fao.org/faostat/en/data/QC>>

de produção e valor de produção. Destes levantamentos, os que estão relacionados a produtos processados não serão utilizados pelo fato de serem produtos da indústria de processamento, que se enquadra na fase "depois da porteira". Dados de valor de produção também não serão utilizados por serem de cunho financeiro. Os dados contidos nos índices de produção não serão utilizados por dizerem respeito aos totais obtidos na outras seções (cultivo, animais vivos, produtos obtidos de animais e etc), dados estes que já serão tirados das outras seções do portal eletrônico da FAO.

Entre os dados relacionados ao cultivo disponibilizados pela FAO encontram-se métricas de produção, área colhida e produtividade de 164 produtos de 1961 a 2016. Apesar de dar a opção de escolha de 164 diferentes produtos, ao selecionar toda a listagem e o Brasil como país desejado, a listagem resultante traz 78 diferentes produtos, apresentados no Apêndice D.

No que se refere a contagem de animais vivos, a FAO lista 19 tipos diferentes de animais entre 1961 a 2016, obtendo resultado para 13 diferentes tipos de rebanhos (Apêndice D).

A produção de produtos primários de origem animal permite a seleção da contagem de animais abatidos, produção e produtividade de 34 diferentes itens, obtendo resultados para 26. Assim como nos outros dados, os filtros permitem selecionar período temporal de 1961 até 2016 e estão representados no Apêndice D.

A ferramenta da FAO sabe identificar, através dos filtros selecionados pelo usuário, qual medida será atribuída a qual produto. Por exemplo, para o produto "cera de abelha" constam somente dados de produção, e não de abate e produtividade.

Diferentemente das outras fontes de dados, a FAO não disponibiliza os dados com divisão estados, sendo somente os dados do Brasil como um todo. Além da disponibilização dos dados acima descritos, a FAO disponibiliza também dados financeiros, de investimento, importação e exportação, relacionados a fertilizantes, inseticidas, entre outros.

4.2 COLETA DOS DADOS

A partir da seleção das fontes de dados (CONAB, EMBRAPA, IBGE e FAO), o segundo objetivo específico indica a geração do modelo de extração automática dos dados de cada uma das bases encontradas. Foi desenvolvido um algoritmo de extração específico para extração em cada uma das fontes.

4.2.1 Coleta de dados CONAB

Conforme levantamento feito na seção anterior, a CONAB disponibiliza 33 planilhas, sendo que 29 serão utilizadas, portanto, o algoritmo deve realizar o *download* de todo este conteúdo, que contempla dados de todos os estados e regiões do Brasil.

Ao analisar o conteúdo do código-fonte da página da CONAB, percebe-se que há *links* para acessar e baixar as planilhas com os dados do portal em quatro páginas, e que estes *links* estão sempre dentro de *tags* com o identificador "catItemAttachments" (Figura 36).

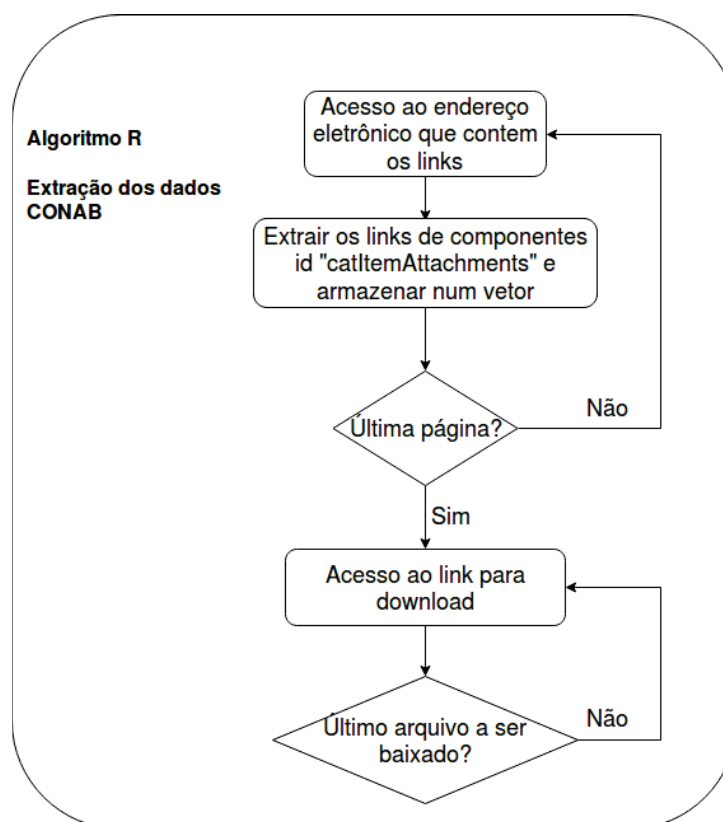
Figura 36 – CÓDIGO FONTE DO PORTAL DA CONAB

```
<div class="row-fluid">
  <!-- Item attachments -->
  <div class="catItemAttachmentsBlock pull-left">
    <ul class="catItemAttachments" style="margin-left: -8px;">
      <li><!-- class="btn" -->
        <a title="Amendoim1aSerieHist.xls" href="/info-agro/safras/serie-historica-das-safras/item/download/22108_84c6fcd6dd218ca7710ad6be20ee54a3">
          <i class="icon-download-alt"></i>
          Amendoim1aSerieHist.xls
          <span>(1170 Downloads)</span>
        </a>
      </li>
    </ul>
  </div>
</div>
```

Fonte: CONAB (2018)

Desta forma, técnicas de *web scraping* possibilitam selecionar o conteúdo da *tag* de id "catItemAttachments", e após isso, do valor do atributo "href" dentro da *tag* "a", e com isso, obter o endereço para acessar e baixar os documentos dispostos pela CONAB. Após a obtenção dos *links*, o *download* do conteúdo deve ser realizado no diretório desejado. O fluxograma de funcionamento do algoritmo de extração de dados da CONAB está disposto na Figura 37.

Figura 37 – FLUXOGRAMA DE EXTRAÇÃO DE DADOS CONAB



Fonte: o Autor (2018)

O algoritmo primeiramente percorre as quatro páginas do portal da CONAB extraindo de cada uma delas os endereços eletrônicos a serem acessados para baixar os dados, e os armazena em um vetor. Após a extração dos *links*, há uma nova iteração, que realiza o acesso e *download* dos dados contidos no vetor gerado no começo do algoritmo. O algoritmo que realiza o procedimento de coleta de dados da CONAB é apresentado no Apêndice E, entre as linhas 501 à 611.

Há de se salientar que caso ocorram mudanças no endereço eletrônico da CONAB onde as planilhas são disponibilizadas, pode comprometer o funcionamento do *script* de coleta de dados, pelo fato do algoritmo ter sido desenvolvido exclusivamente para funcionamento na retirada de parâmetros das *tags* com identificador "catItemAttachments".

4.2.2 Coleta de dados EMBRAPA

Dentro do portal da EMBRAPA é possível tanto visualizar os dados diretamente no navegador quanto realizar o *download* de arquivos com dados em formatos como .PDF, .XLS, .XLSX entre outros, porém, o Agropensa é um aplicativo que utiliza a

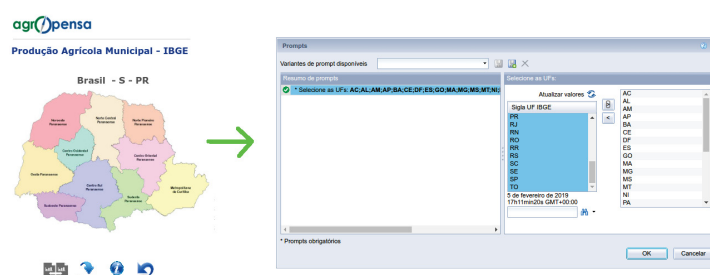
tecnologia Flash, o que, neste trabalho, acabou impossibilitando a coleta automática tanto via *web scraping* quanto a realização do *download* dos arquivos.

Trabalhos como de Pikor e Piotrowski (2011), Barcaroli et al. (2014), e mais recentemente por Barcaroli et al. (2016) apontam a dificuldade de se obter dados de forma automática de aplicações construídas utilizando a tecnologia Flash.

Sendo assim, optou-se, neste caso, pela coleta dos dados de forma manual, realizando o *download* dos arquivos com os dados em formato .XLSX para que pudessem ser submetidos a limpeza dos mesmos através de um algoritmo. Foi realizada a coleta manual de um total de 121 arquivos .CSV, divididos em 79 com dados de agricultura e 42 de pecuária. A coleta dos arquivos ocorreu no dia 07/01/2019.

Para a coleta dos dados, foi necessário acessar a página inicial do AGROPENSA⁷, clicar em um dos estados até que aparecesse o botão "comparar" no canto inferior esquerdo e, após isso, selecionar todos os estados do Brasil (Figura 38).

Figura 38 – SELEÇÃO DOS ESTADOS PARA EXTRAÇÃO DE DADOS DA EMBRAPA

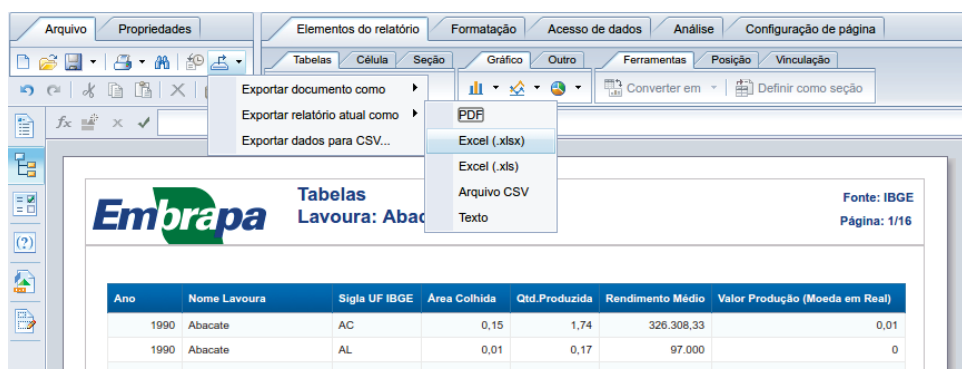


Fonte: o Autor (2018)

Após a seleção dos estados e do produto que se deseja obter os dados, o AGROPENSA disponibiliza uma tela com diversos gráficos. A aba chamada "Tabelas" contempla os dados, onde pode ser realizado o *download* do arquivo com os dados no formato desejado (Figura 39).

⁷ <<https://www.embrapa.br/agropensa/bases-de-dados>>

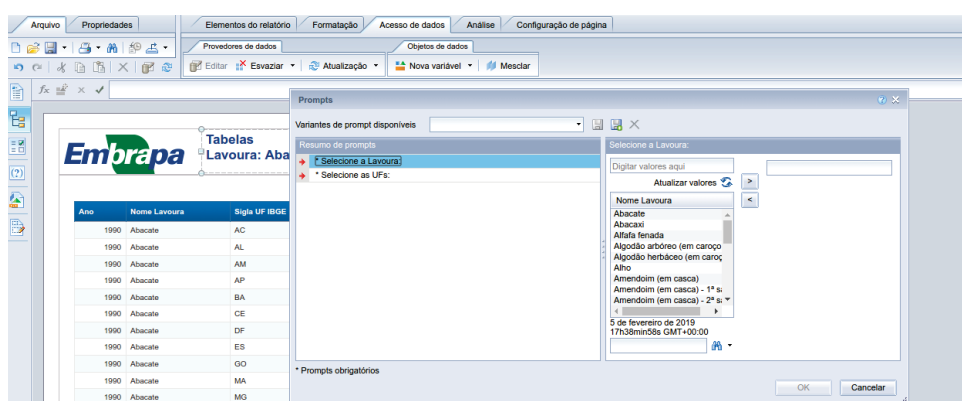
Figura 39 – DOWNLOAD DO ARQUIVO COM DADOS DA EMBRAPA



Fonte: o Autor (2018)

Após o *download* do arquivo, é possível, dentro da opção "Acesso aos dados" no centro superior da tela, realizar nova pesquisa. Ao clicar no botão "Atualização" é possível redefinir os parâmetros (Figura 40) para geração de nova tabela que será baixada.

Figura 40 – REDEFINIÇÃO DE PARÂMETROS PARA OBTENÇÃO DE DADOS DA EMBRAPA



Fonte: o Autor (2018)

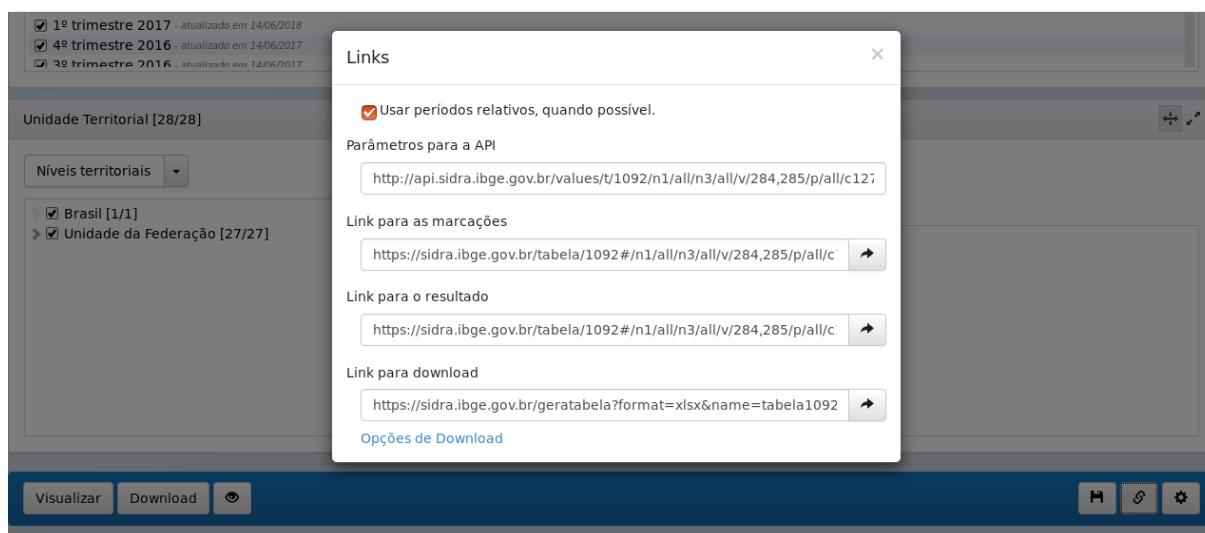
Este procedimento foi repetido até que fossem baixadas todas as planilhas com dados de todos os produtos indicados no Apêndice B.

4.2.3 Coleta de dados IBGE

Dentro de cada uma das abas, o SIDRA possibilita que o usuário selecione parâmetros para seleção dos dados que deseja, dentre estes parâmetros, encontram-se: o tipo do indicador, referência temporal, trimestre, localização, entre outros. Além de possibilitar a visualização e *download* direta dos dados conforme os filtros selecionados,

o IBGE disponibiliza também *links* que podem ser utilizados em alguma *Application Programming Interface* (API), e até mesmo o endereço eletrônico para download direto dos dados (Figura 41).

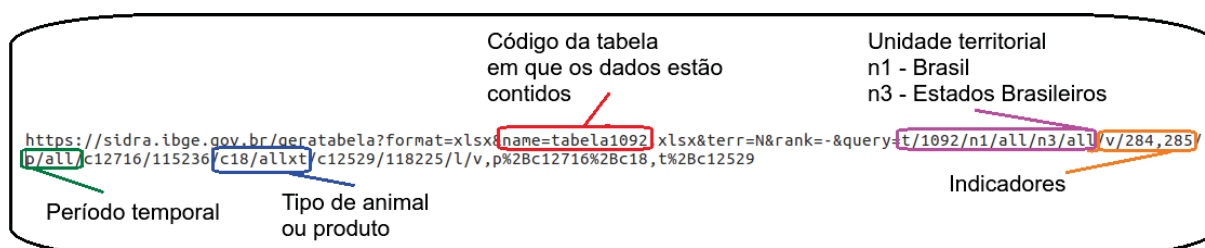
Figura 41 – DISPONIBILIZAÇÃO DOS DADOS DA PLATAFORMA SIDRA - IBGE



Fonte: IBGE (2018)

A partir do endereço disponibilizado (Figura 42) para *download*, é possível notar a presença dos parâmetros selecionados na ferramenta para a construção do mesmo, o que tornou possível automatizar a coleta de todas as planilhas de dados disponibilizadas pelo IBGE.

Figura 42 – ENDEREÇO ELETRÔNICO PARA OBTENÇÃO DOS DADOS DO IBGE

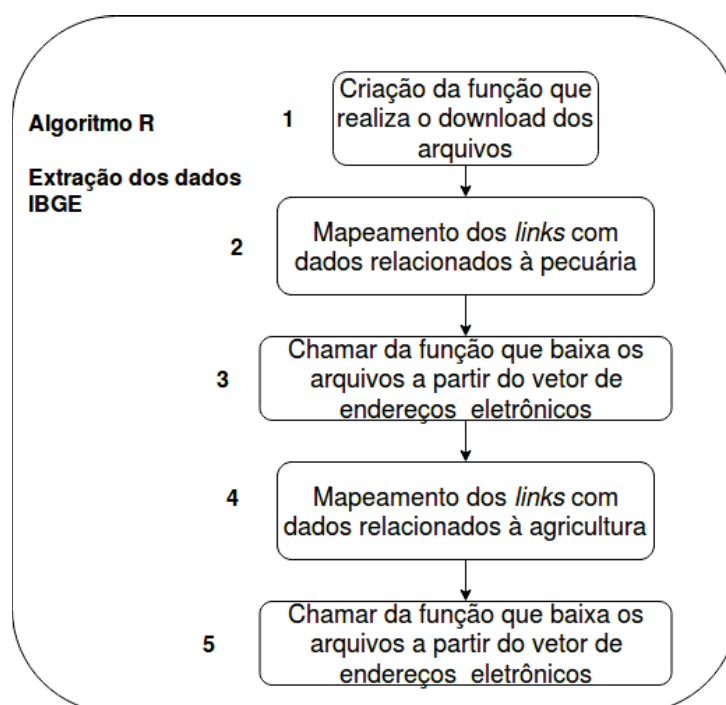


Fonte: o Autor (2018)

Os parâmetros variam de acordo com cada tipo de dado e produto. Por exemplo, a tabela 1092 identificada na figura acima diz respeito à tabela com dados relacionados ao abate de bovinos, enquanto o identificador de tabela 6588 contém dados de produtos agrícolas.

Os valores do código dos indicadores e do tipo de produto ou animal também variam de acordo com a necessidade do usuário. Sendo assim, é necessário que o algoritmo que realize a obtenção dos dados contemple todos os endereços eletrônicos para *download* dos dados, contemplando maior amplitude temporal e territorial possível. A Figura 43

Figura 43 – FLUXOGRAMA DE EXTRAÇÃO DE DADOS IBGE



Fonte: o Autor (2018)

Conforme o fluxograma acima, o algoritmo foi construído para coletar os dados em etapas de acordo com a ordem das abas disponibilizadas pela plataforma SIDRA: abate, produção de leite, ovos e agrícola. O *script* completo de coleta de dados do IBGE é apresentado no Apêndice E e realiza as seguintes ações:

- Passo 1 - a primeira etapa diz respeito a função que realiza o *download* dos arquivos mapeados nos passos seguintes (linhas 375 à 382 do Apêndice E);
- Passo 2 - realizar o mapeamento dos links para obtenção dos arquivos (linhas 384 à 427 do Apêndice E);
- Passo 3 - executar o *download* dos arquivos relacionados a pecuária chamando a função do Passo 1 (linhas 410, 420 e 430 do Apêndice E).
- Passo 4 - realizar o mapeamento como no Passo 2, só que desta vez com arquivos relacionados à agricultura (linhas 438 à 462 do Apêndice E);

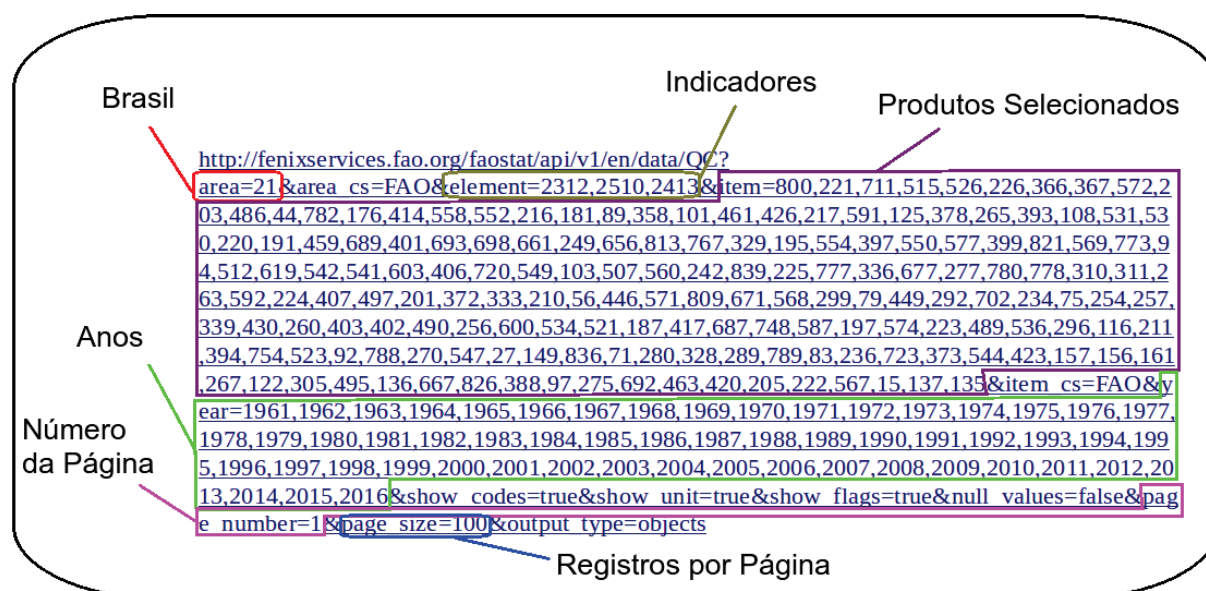
- Passo 5 - invocar a função que realiza o download de arquivos (linha 465 do Apêndice E).

Diferente da extração dos dados da CONAB, a alteração da parte visual portal eletrônico do IBGE não necessariamente acarretaria em problemas para a extração automática de seus dados. Neste caso, o que pode trazer problemas seria a alteração da composição do endereço eletrônico para obtenção dos dados, já que estes estão indicados dentro do código-fonte utilizado para extração dos dados. Também há a restrição da não captação de dados de novos produtos que possam ser inseridos no portal do IBGE, pois os *links* para acessar os dados estão inseridos dentro do algoritmo, não sendo pegos diretamente do sítio eletrônico.

4.2.4 Coleta de dados FAO

A FAO⁸ possibilita, além do *download* de planilhas, realizar uma requisição para que o portal retorne um JSON com os dados. Ao se observar a *Uniform Resource Locator* (URL), nota-se a presença dos parâmetros que trarão os dados selecionados na pagina anterior. (Figura 44).

Figura 44 – ENDEREÇO ELETRÔNICO DA FAO PARA ACESSO AO JSON COM DADOS



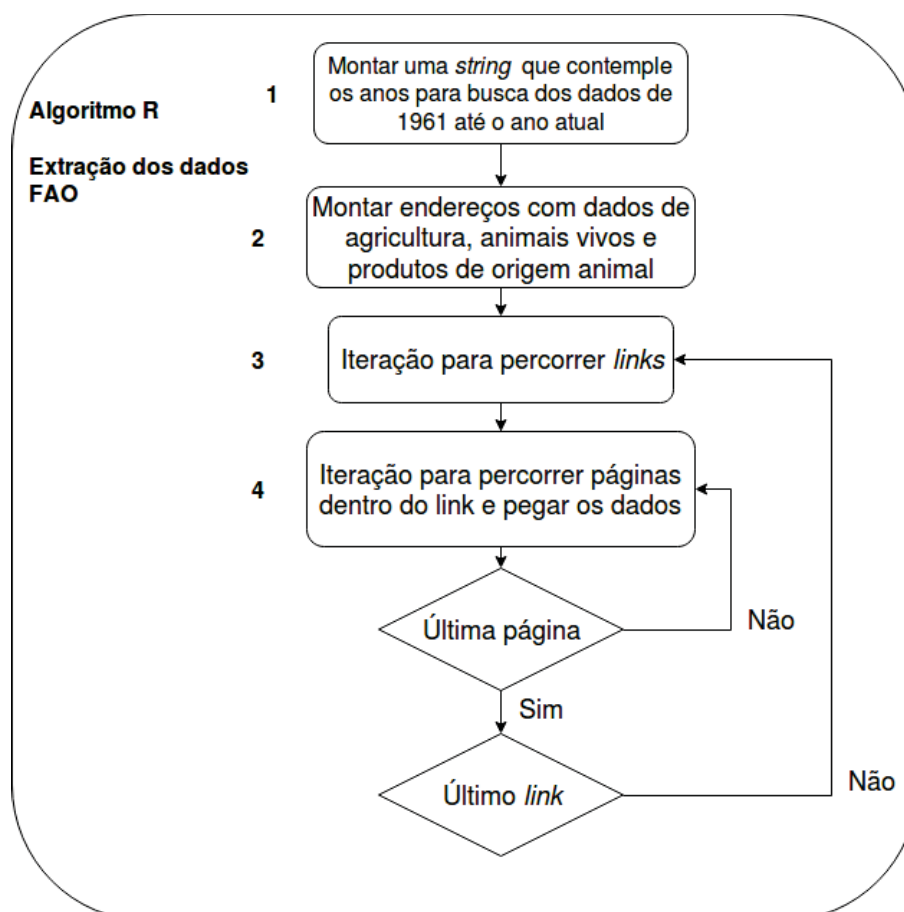
Fonte: o Autor (2018)

A partir do entendimento da URL, a construção do algoritmo baseou-se, primeiramente, na montagem dos parâmetros para realizar as requisições no portal e coletar

⁸ <<http://www.fao.org/faostat/en/#data>>

os dados e após isso em dois passos principais de iterações para extração dos dados (Figura 45).

Figura 45 – FLUXOGRAMA DE EXTRAÇÃO DE DADOS FAO



Fonte: o Autor (2018)

Na primeira parte do fluxo do algoritmo, são indicados como parâmetros sempre dados de 1961 até o ano atual, para, no caso dos dados serem atualizados, que sejam também recolhidos pelo *script* (linhas 1.019 à 1.027 do Apêndice E). A segunda parte do *script* diz respeito a montagem do endereço principal a ser acessado para a extração dos dados (linhas 1.030 à 1.054 do Apêndice E).

A partir da terceira etapa, duas iterações ocorrem: a primeira delas busca acessar todos os endereços principais mapeados no passo anterior (linhas 1.058 à 1.148 do Apêndice E). A próxima iteração, quarto passo, ocorre dentro da primeira, realizando o acesso a página para extração dos dados, sempre começando pela página 1 e é realizada enquanto existem dados dentro das páginas (linhas 1.061 à 1.146 do Apêndice E).

Alguns endereços principais contêm mais de 100 registros, e por isso são

separados em várias páginas. O algoritmo percorre todas estas páginas que possuem dados. Ao verificar que uma página possui menos do que o número padrão de registros (100), significa que esta é a última página, parando a execução da iteração de extração dos dados (linha 1.066 à 1.068 do Apêndice E).

A coleta de dados da FAO contempla os mesmos riscos e limites apontados na coleta de dados do IBGE. O algoritmo não consegue coletar dados, automaticamente, de novos produtos caso sejam incluídos pela FAO sem uma manutenção no algoritmo, e a alteração da estrutura do endereço de acesso aos dados pode fazer com que o *script* não funcione corretamente.

4.3 TRATAMENTO DOS DADOS

Após a construção dos algoritmos responsáveis pela coleta dos dados a partir das quatro fontes selecionadas, se faz necessário realizar o tratamento dos dados, com intuito de torná-los uniformes independentemente de sua origem, seguindo o padrão indicado na seção de Tratamento dos Dados dentro do capítulo de Materiais e Métodos (3.3.3, página 86), e por consequência, realizar o terceiro objetivo específico.

Assim como realizado na coleta de dados, foi construído um algoritmo específico para a limpeza dos dados de cada uma das fontes de dados diferentes que foram reunidos em um único arquivo (Apêndice E). O funcionamento de todos os algoritmos de limpeza depende de que o padrão dos dados e arquivos coletados na seção anterior sejam mantidos. A alteração da forma de disponibilização dos dados ou da estrutura do arquivo com os dados pode comprometer o funcionamento do algoritmo de limpeza dos como um todo.

4.3.1 Tratamento dos dados CONAB

Os arquivos que contêm os dados da CONAB são disponibilizados em formato .XLS, sendo que cada um deles contempla dados de um tipo específico de produto, conforme demonstrado na Figura 46.

Figura 46 – ESTRUTURA DA PLANILHA COM DADOS DA CONAB

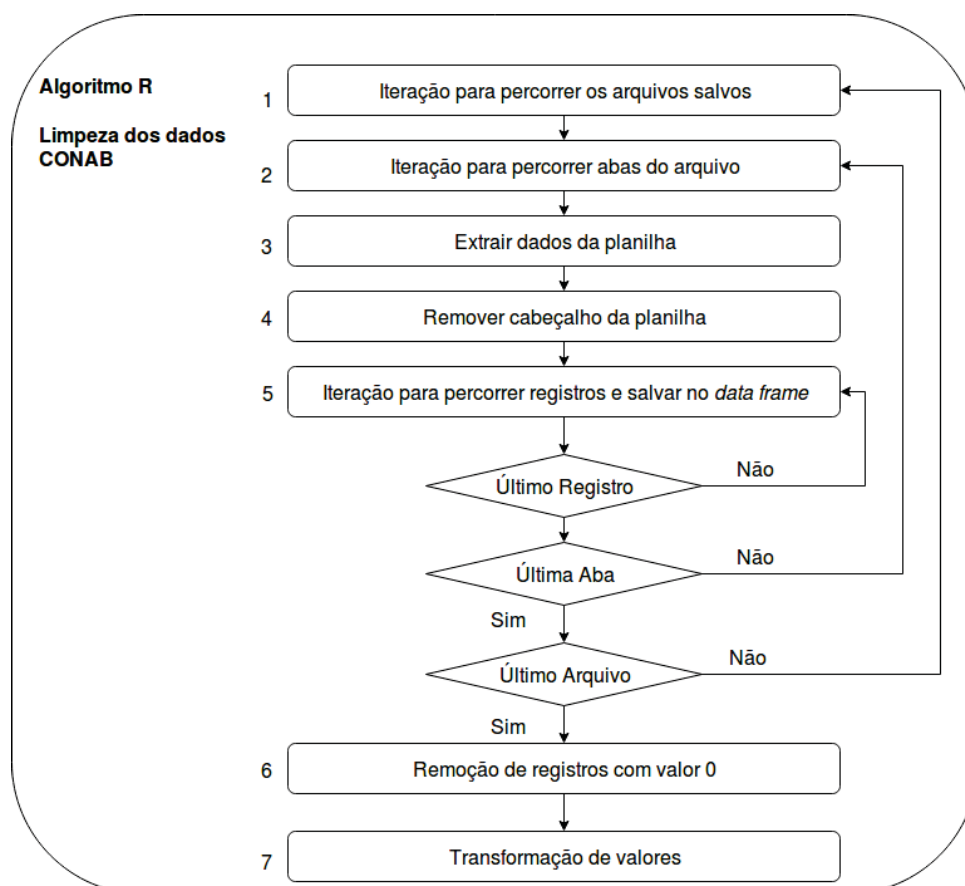
	A	U	V	W	X	Y	Z	AA	AB	AC	AD
1											
2	Produto	SOJA	BRASIL								
3		Historica d	Area Plantada								
4	Anos	Safras 1976/77 a 2018/19									
5		Em mil hectares									
6	REGIÃO/UF	1995/96	1996/97	1997/98	1998/99	1999/00	2000/01	2001/02	2002/03	2003/04	2004/05
7	NORTE	6,7	25,2	47,4	50,4	70,6	91,7	141,1	209,7	352,4	521,9
8	RR	1,8	-	-	-	-	-	3,5	3,0	12,0	20,0
9	RO	-	3,3	4,7	8,7	11,8	25,0	28,6	41,0	59,5	74,4
10	AC	-	-	-	-	-	-	-	-	-	-
11	AM	-	-	-	-	-	-	1,1	2,1	2,1	2,8
12	AP	-	-	-	-	-	-	-	-	-	-
13	PA	-	-	2,6	1,6	2,3	0,7	2,9	15,5	35,2	69,0
14	TO	4,9	21,9	40,1	40,1	56,5	66,0	105,0	148,1	243,6	355,7
15	NORDESTE	532,3	593,9	728,9	772,8	851,0	962,6	1.125,1	1.240,7	1.323,3	1.442,1
16	MA	89,1	120,0	144,0	162,7	175,7	210,0	238,3	274,0	342,5	375,0
17	PI	10,2	17,9	28,6	29,9	40,0	62,0	86,8	116,3	159,3	197,1
18	CE	-	-	-	-	-	-	-	-	-	-
19	RN	-	-	-	-	-	-	-	-	-	-
20	PB	-	-	-	-	-	-	-	-	-	-
21	PE	-	-	-	-	-	-	-	-	-	-
22	AL	-	-	-	-	-	-	-	-	-	-
23	SE	-	-	-	-	-	-	-	-	-	-
24	BA	433,0	456,0	556,3	580,2	635,3	690,6	800,0	850,4	821,5	870,0
25	CENTRO-OESTE	3.694,7	3.983,8	5.060,2	4.955,1	5.499,4	5.759,5	6.985,0	8.048,4	9.659,3	10.857,0
26	MT	1.905,2	2.095,7	2.600,0	2.548,0	2.904,7	3.120,0	3.853,2	4.419,6	5.240,5	6.105,2
27	MS	845,4	862,3	1.086,5	1.053,9	1.106,6	1.064,5	1.192,2	1.415,1	1.797,2	2.030,8
28	GO	909,4	991,2	1.338,1	1.324,7	1.454,5	1.540,0	1.901,9	2.170,5	2.572,0	2.662,0
29	DF	34,7	34,6	35,6	28,5	33,6	35,0	37,7	43,2	49,6	59,0
30	SUDESTE	1.091,6	1.097,6	1.131,1	1.097,6	1.152,9	1.172,0	1.296,7	1.488,9	1.826,9	1.891,6
31	MG	528,0	522,7	601,1	577,1	594,4	642,0	719,0	873,6	1.065,8	1.119,1
32	ES	-	-	-	-	-	-	-	-	-	-
33	RJ	-	-	-	-	-	-	-	-	-	-
34	SP	563,6	574,9	530,0	520,5	558,5	530,0	577,7	615,3	761,1	772,5
35	SUL	5.337,9	5.680,8	6.190,3	6.119,3	6.049,0	5.984,0	6.838,3	7.487,1	8.213,9	8.588,5
36	PR	2.311,5	2.496,4	2.820,0	2.769,2	2.832,9	2.818,0	3.291,4	3.637,6	3.935,9	4.148,4
37	SC	222,4	240,2	220,0	215,6	207,0	196,0	241,3	255,8	307,0	350,0
38	RS	2.804,0	2.944,2	3.150,3	3.134,5	3.009,1	2.970,0	3.305,6	3.593,7	3.971,0	4.090,1
39	NORTE/NORDESTE	539,0	619,1	776,3	823,2	921,6	1.054,3	1.266,2	1.450,4	1.675,7	1.964,0
40	CENTRO-SUL	10.124,2	10.762,2	12.381,6	12.172,0	12.701,3	12.915,5	15.120,0	17.024,4	19.700,1	21.337,1
41	BRASIL	10.663,2	11.381,3	13.157,9	12.995,2	13.622,9	13.969,8	16.386,2	18.474,8	21.375,8	23.301,1
42	Legenda: (*) Estimativa										
43	Fonte: Conab										
44											
45											

Fonte: o Autor (2018)

A estrutura das planilhas disponibilizadas pela CONAB possibilitou a construção de um algoritmo de que retirasse os dados marcados na imagem acima de maneira automática. Logo no topo da planilha é possível encontrar o nome do produto, destacado em verde. Dados do indicador e da unidade utilizada são encontrados logo a seguir, demarcados nas cores vermelho e rosa respectivamente.

Por fim, a estrutura apresenta uma tabela onde cada linha representa um local, que podem ser um estado, uma região ou o Brasil como um todo. As colunas da tabela contemplam os anos das medições. Desta forma, foi possível um construir um algoritmo que, seguindo o fluxo demonstrado na Figura 47, realizasse a limpeza dos dados no padrão definido no capítulo de Materiais e Métodos do presente trabalho através de sete passos.

Figura 47 – FLUXOGRAMA DE LIMPEZA DE DADOS CONAB



Fonte: o Autor (2018)

- Passo 1 - a primeira etapa do algoritmo compreende numa iteração para percorrer todos os arquivos obtidos na seção Coleta de dados CONAB (4.2.1, p.103). A iteração encontra-se entre as linhas 528 e 592 do Apêndice E;
- Passo 2 - uma segunda iteração ocorre dentro da primeira, e é responsável por percorrer cada uma das abas do arquivo (linhas 540 à 590 do Apêndice E);
- Passo 3 - são extraídos das primeiras linhas do arquivo: nome do produto, indicador e unidade (linhas 546 à 555 do Apêndice E);
- Passo 4 - as linhas que já tiveram seus dados extraídos são removidas, de modo que reste apenas a tabela com os dados (linhas 556 à 560 do Apêndice E);
- Passo 5 - entre as linhas 573 e 588 do Apêndice E são percorridos registros linha por linha, coluna por coluna da tabela para retirada dos dados e inserção num *data frame* com o formato padrão definido para o presente trabalho;
- Passo 6 - são removidos registros com valor zero ou nulo (linhas 597 e 598 do Apêndice E);

- Passo 7 - o último passo contempla a remoção de dados que tem indicadores diferente dos selecionados para o presente trabalho, como por exemplo área de renovação, área de expansão e etc. Também são retirados dados que na coluna "Estado" contemplam duas regiões, como por exemplo "Norte/Nordeste" (linhas 601 à 606 do Apêndice E).

Caso algum problema ocorra durante a coleta e limpeza dos dados da CONAB, é gerado um arquivo formato .TXT com o nome "erro limpeza CONAB data hora" que indica qual o erro ocorrido durante a tarefa diária (linha 610 do apêndice E).

4.3.2 Tratamento dos dados EMBRAPA

As planilhas com dados disponibilizados pela EMBRAPA possuem uma estrutura diferente dos demais (Figura 48), caracterizando-se principalmente pela falta da indicação de qual produto os dados se relacionam. Para resolver isso, os arquivos foram salvos com o nome do produto, como por exemplo: Cebola.xlsx, Soja.xlsx e etc.

Figura 48 – ESTRUTURA DA PLANILHA COM DADOS DA EMBRAPA

Ano	UF	Área Colhida(em Milhares de Hectares)	Qtd.Produzida(em Mil Toneladas)	Rendimento Médio(em Kilos por hectares)	Valor Produção (em Milhares de R\$)
	AC				
	AL				
	AM				
	AP				
	BA	5,096	68,485	575100	0,64994
	CE	0,011	0,017	6500	0,0003236364
	DF	0,005	0,058	11600	0,0006152727
	ES	0,076	0,458	38000	0,0091189091
	GO				
	MA				
	MG	0,781	5,447	648494,99722327	0,0432516364
	MS	0,028	0,241	27000	0,0021814545
	MT				
	PA				
	PB	0,018	0,073	22500	0,0006698182
	PE	2,75	34,302	229000	0,3901810909
	PI	0,048	0,227	55833,3333333333	0,0053618182
	PR	5,543	44,613	887864,102564103	0,2487298182
	RJ				
	RN	0,053	0,931	46333,3333333333	0,0125763636
	RO				
	RR				
	RS	17,271	131,647	2125492,80413734	0,4815312727
	SC	27,278	306,525	1032100	1,2626461818
	SE	0,008	0,038	9000	0,0005527273
	SP	15,68	275,997	763387,287207305	3,7638236364
	TO				
199	Total:	74,646	869,067	6478205,85779868	6,8715036364
	AC				
	AL				
	AM				
	AP				

Fonte: o Autor (2018)

Na imagem é possível identificar que os dados os anos das medições são indicados no canto esquerdo do documento, seguido do estado e dos resultados

encontrados. Além dos dados de área colhida e produção, encontram-se também dados de produtividade e valor de produção, ambos não utilizados.

Os dados relacionados ao valor de produção serão descartados por não fazerem parte do escopo da presente pesquisa, já os dados de produtividade não foram utilizados pelo fato de, por algumas vezes, serem calculados errados pela ferramenta Agropensa (Figura 49).

Figura 49 – PROBLEMA DE CÁLCULO DA PRODUTIVIDADE NOS DADOS DA EMBRAPA

PI	0,048	0,227	55833,3333333333
PR	5,543	44,613	887864,102564103

Área Colhida
Produção
Produtividade

Fonte: o Autor (2018)

O cálculo da produtividade se dá através da multiplicação por 1000 da divisão entre o total produzido e a área colhida ($1000 \times (\text{produção} / \text{área colhida})$). Sendo assim, o resultado para o estado do Piauí seria de aproximadamente 4.729 quilogramas por hectare ($1000 \times (0,227 / 0,048)$), e não 55.833 como apresentado no arquivo com os dados. Já para o estado do Paraná, também apresentado na imagem, o valor correto seria 8.048,52 ao invés de 887.864,10. Desta forma, optou-se pela própria ferramenta realizar o cálculo da produtividade a partir dos dados disponibilizados pela EMBRAPA.

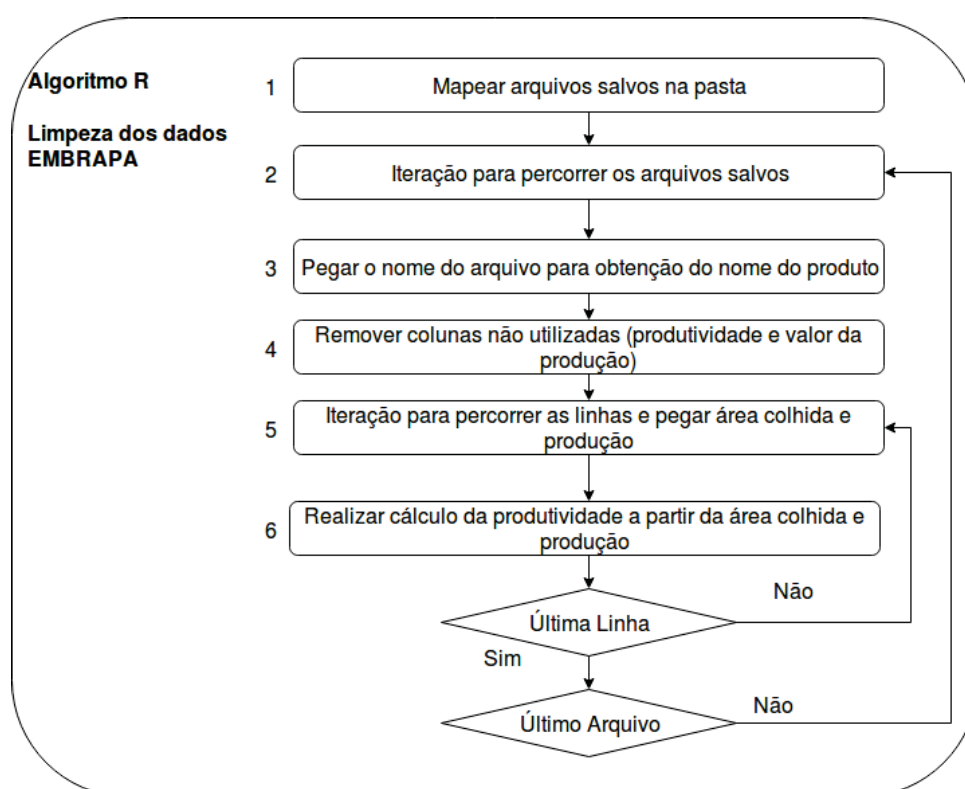
A limpeza dos dados da EMBRAPA obtidos através da obtenção manual dos arquivos é realizada através de seis passos principais (Figura 50):

- Passo 1 - verificar arquivos existentes dentro da pasta de agricultura e pecuária da EMBRAPA e inserir o nome dos mesmos em uma lista (linhas 616 e 617 do Apêndice E);
- Passo 2 - inicia-se a primeira iteração, onde são percorridos todos os arquivos contidos na lista adquirida no Passo 1 (linhas 797 à 865 para dados de agricultura e 867 à 1.010 para pecuária; todas no Apêndice E);
- Passo 3 - como para alguns arquivos não consta o nome do produto entre os dados, o algoritmo pega o nome do arquivo, que é o nome do produto. Por exemplo, o nome do arquivo "Cebola.xlsx" torna-se somente "Cebola" (linhas 811 e 812 para dados de agricultura, 955 e 956 para dados de ovinos tosquiados e 960 e 961 para vacas ordenhadas e leite). Para dados de aquicultura, rebanho abate e produtos de origem animal, o nome dos produtos se encontra dentro do

arquivo entre os dados, e são extraídos nas linhas 890 à 894 no caso de dados de aquicultura, rebanhos e produtos de origem animal, 931 e 942 para dados de abate;

- Passo 4 - remoção das colunas não utilizadas conforme indicado nesta seção (linhas 804 e 805 do Apêndice E para dados de agricultura; 873 e 874 para dados de aquicultura, rebanhos e produtos de origem animal; 921 e 922 para abate);
- Passo 5 - inicia-se a iteração para coletar os dados linha a linha e inserir no *data frame* (no Apêndice E linhas 816 à 860 para dados de agricultura; 886 à 917 para aquicultura, rebanhos e produtos de origem animal; 924 à 947 para abate e 968 à 1.004 para ovinos tosquiados, vacas ordenhadas e leite);
- Passo 6 - para os dados de agricultura, assim que são coletados os dados de área colhida e produção, calcula-se a produtividade (linha 831 do Apêndice E).

Figura 50 – FLUXOGRAMA DE LIMPEZA DE DADOS EMBRAPA



Fonte: o Autor (2018)

Como nos dados do IBGE não consta os totais por região, foi criada uma função genérica (linhas 620 à 795 do Apêndice E) para realizar o cálculo ano a ano de cada produto por região. Para realizar o cálculo a função precisa receber como parâmetros o

data frame com os dados, o nome do indicador a ser calculado, o nome do produto e da unidade. A chamada da função é realizada dentro da iteração do Passo 2, logo após a realização da iteração para inserir os dados no *data frame* (Passo 5). A invocação da função para dados de agricultura acontece nas linhas 861, 862 e 863; 911 à 917 para aquicultura, rebanhos e produtos de origem animal; 949 à 952 para abate e 1.005 à 1.007 para ovinos tosquiados, vacas ordenhas e leite.

Para o controle de erros na limpeza dos dados da EMBRAPA, caso ocorram, a linha 1.013 do Apêndice E escreve um arquivo .TXT informando sobre o erro e o salva na pasta onde encontram-se as pastas com arquivos a serem salvos.

4.3.3 Tratamento dos dados IBGE

Os dados obtidos da plataforma SIDRA do IBGE são obtidos dentro de arquivos de planilha formato .XLSX na estrutura demonstrada na Figura 51.

Figura 51 – ESTRUTURA DA PLANILHA COM DADOS DO IBGE

	A	B	C	D	E	F	G	H	I	J	K
1	Tabela 6588 - Série histórica da estimativa anual da área plantada, área colhida, produção e rendimento médio dos produtos das lavouras										
2	Variável	Área plantada (Hectares)	Indicador								
3		Mês x Produto das lavouras									
4	Mês/Ano	setembro	outubro	novembro	dezembro	janeiro	fevereiro	março	abril	maio	junho
5	Brasil e Unidade da Federação	1.17 Trigo	1.17 Trigo	1.17 Trigo	1.17 Trigo	1.17 Trigo	1.17 Trigo	1.17 Trigo	1.17 Trigo	1.17 Trigo	1.17 Trigo
6	Brasil	1761440	1765644	1757764	1762618	1753930	1753930	1753930	1753863	1778232	1778861
7	Rondônia	-	-	-	-	-	-	-	-	-	-
8	Acre	-	-	-	-	-	-	-	-	-	-
9	Amazonas	-	-	-	-	-	-	-	-	-	-
10	Roraima	-	-	-	-	-	-	-	-	-	-
11	Pará	-	-	-	-	-	-	-	-	-	-
12	Amapá	-	-	-	-	-	-	-	-	-	-
13	Tocantins	-	-	-	-	-	-	-	-	-	-
14	Maranhão	-	-	-	-	-	-	-	-	-	-
15	Piauí	-	-	-	-	-	-	-	-	-	-
16	Ceará	-	-	-	-	-	-	-	-	-	-
17	Rio Grande do Norte	-	-	-	-	-	-	-	-	-	-
18	Paraíba	-	-	-	-	-	-	-	-	-	-
19	Pernambuco	-	-	-	-	-	-	-	-	-	-
20	Alagoas	-	-	-	-	-	-	-	-	-	-
21	Sergipe	-	-	-	-	-	-	-	-	-	-
22	Bahia	-	-	-	-	-	-	-	-	-	-
23	Minas Gerais	10646	12800	12800	12864	10646	10646	10646	9890	11034	10001
24	Espírito Santo	-	-	-	-	-	-	-	-	-	-
25	Rio de Janeiro	-	-	-	-	-	-	-	-	-	-
26	São Paulo	48900	48900	48900	48900	48900	48900	48900	48900	48900	48900
27	Paraná	886194	886602	878722	878722	878722	878722	878722	900436	852340	847592
28	Santa Catarina	59250	60364	60364	62006	59250	59250	59250	59250	59250	59250
29	Rio Grande do Sul	695471	696188	696188	699336	695471	695471	695471	695471	766792	772942
30	Mato Grosso do Sul	50599	50410	50410	50410	50561	50561	50561	29536	29536	29791
31	Mato Grosso	-	-	-	-	-	-	-	-	-	-
32	Goiás	10380	10380	10380	10380	10380	10380	10380	10380	10380	10380
33	Distrito Federal	-	-	-	-	-	-	-	-	-	-
34	Fonte: IBGE - Levantamento Sistemático da Produção Agrícola										
		Área plantada (Hectares)	Área colhida (Hectares)	Produção (Toneladas)	Rendimento médio (Quilograma)	Notas	Abas com dados de diferentes indicadores				

Fonte: o Autor (2018)

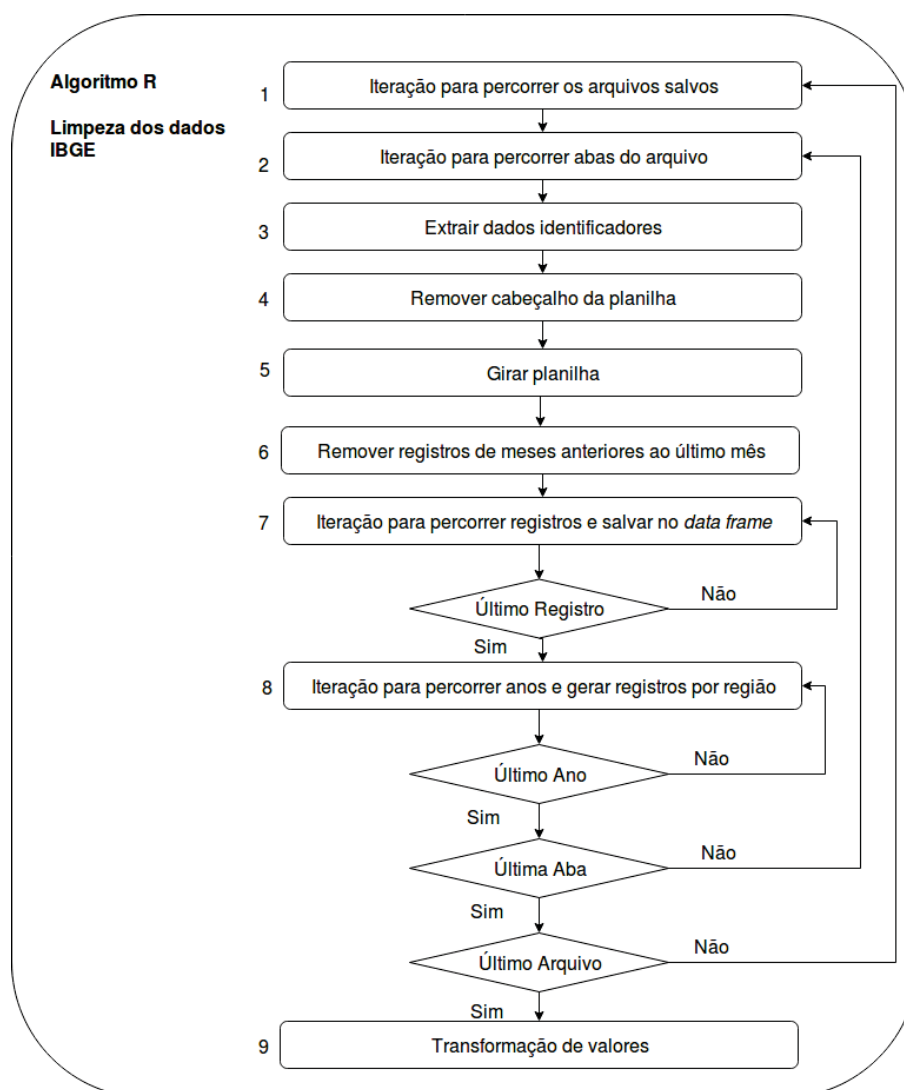
A figura mostra os principais locais para obtenção dos dados. No topo da imagem, destacado em vermelho, é possível visualizar o nome do indicador, logo abaixo, em cinza e verde respectivamente, é possível verificar o período temporal da medição dos dados e o produto medido. Abaixo destas informações encontram-se os estados medidos, e logo a sua direita os valores resultantes das medições. Por fim, na parte inferior da página encontram-se diversas abas, cada uma delas com dados de

um indicador específico. A imagem mostra informações sobre trigo, na aba com dados relacionados ao indicador Área Plantada.

Diferentemente das outras fontes de dados, o IBGE lança previsões mensais do valor de seus indicadores aos invés de lançar apenas uma vez, anualmente. Com isso, os dados vão se ajustando conforme o decorrer do ano, chegando ao valor correto sempre em Dezembro (no caso do ano já ter sido encerrado). Caso o ano não tenha sido encerrado, o valor mais correto é sempre do mês mais próximo a Dezembro. Com isso, o campo "valor" a ser pego é o de Dezembro ou, no caso de não existir, o mais perto possível.

Para realizar a limpeza dos dados do IBGE e deixá-los no formato definido na seção Tratamento dos dados (3.3.3, p.86), o algoritmo responsável pela ação realiza 10 passos, condensando quatro iterações para a realização da limpeza (Figura 52). Os passos abaixo descritos encontram-se todos no algoritmo localizado no Apêndice E.

Figura 52 – FLUXOGRAMA DE LIMPEZA DE DADOS IBGE



Fonte: o Autor (2018)

- Passo 1 - inicia-se a primeira iteração (linha 51 à 208 para dados de agricultura e 211 à 374 para pecuária) através de uma função onde são percorridos todos os arquivos baixados pela extração de dados apresentada na seção Coleta de dados IBGE (4.2.3, p.106);
- Passo 2 - começo da segunda iteração, que percorre cada uma das abas contida em cada um dos arquivos baixados (linhas 53 à 206 para dados de agricultura e 213 à 372 para pecuária);
- Passo 3 - são extraídos os dados do nome do indicador e do produto, contidos sempre nas primeiras linhas do arquivo (linhas 63 à 74 para agricultura e 227 à 241 para pecuária);

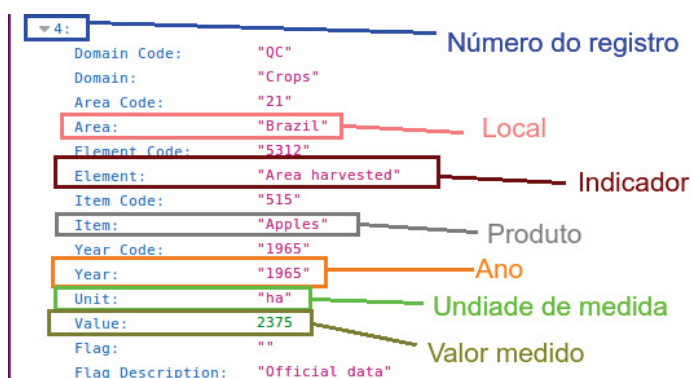
- Passo 4 - primeiramente é carregada a planilha dentro da do R, após, ocorre remoção da primeira linha do arquivo; que não contém dados a serem utilizados (linha 77 para agricultura e 244 para pecuária);
- Passo 5 - como a planilha padrão não segue o conceito de *tidy data*, é realizada uma transposição da tabela, para que, em cada linha, fique apenas uma observação, seguindo assim o conceito utilizado neste trabalho (linhas 83 à 87 para agricultura e 249 à 253 para pecuária);
- Passo 6 - são removidos registros de meses anteriores ao último registrado em cada um dos anos. A descrição de Mês/Ano é alterada apenas para ano, para seguir o padrão selecionado (linhas 86 à 97 para pecuária e 254 à 259 para agricultura);
- Passo 7 - após a planilha estar em formato correto e com dados de anos filtrados, ocorre a terceira iteração, que realiza a leitura das linhas da planilha formatada e os insere no *data frame* que armazena os dados do IBGE (entre as linhas 102 e 111 para agricultura e 262 e 276 para pecuária);
- Passo 8 - após finalizar a inserção dos dados ocorre outra iteração, responsável por gerar o registro ano a ano por região. Por exemplo, para cada um dos anos são somados valores obtidos dos estados Paraná, Santa Catarina e Rio Grande do Sul, e assim é gerado o registro Sul. Essa ação é realizada para cada uma das cinco regiões do Brasil (linhas 118 à 206 para agricultura e 279 à 372 para pecuária);
- Passo 9 - após a limpeza dos dados tanto de pecuária como de agricultura, é realizada a transformação de dados que estão em medidas diferentes das selecionadas na seção de Tratamento dos dados e a remoção de dados com valor zero (linhas 470 à 491).

Assim como na limpeza dos dados da CONAB e EMBRAPA, caso algum erro aconteça durante a coleta e limpeza dos dados do IBGE, um arquivo é gerado com o erro ocorrido (linha 495 do Apêndice E)

4.3.4 Tratamento dos dados FAO

Diferentemente das outras fontes de dados, os dados da FAO foram coletados via JSON que possui a estrutura apresentada da Figura 53:

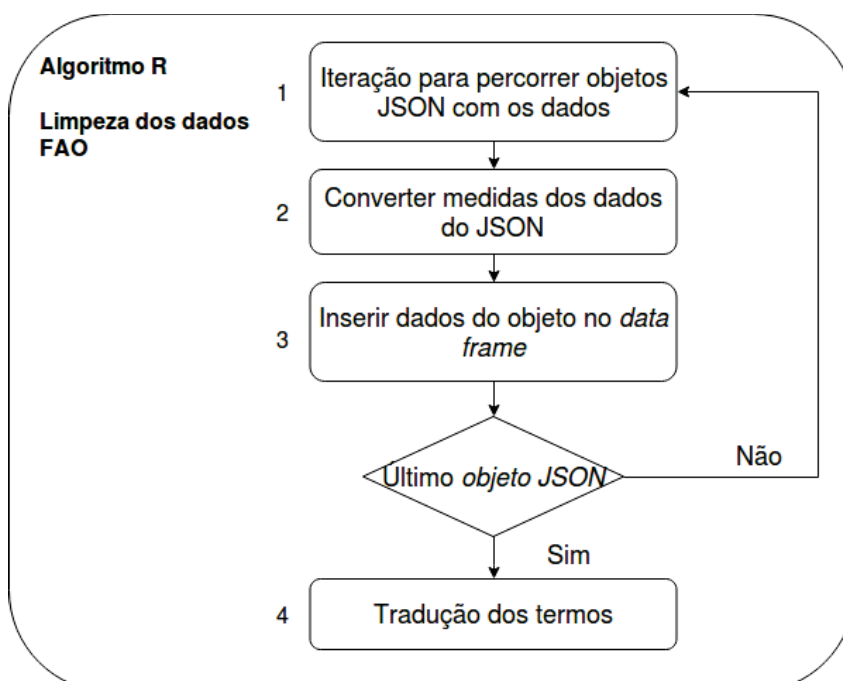
Figura 53 – ESTRUTURA DO JSON DA FAO



Fonte: o Autor (2018)

O topo da imagem mostra, o número do registro, e dentro dele, encontram-se os dados de área, indicador, produto, ano, unidade e valor com os seguintes nomes. Desta forma, o algoritmo responsável pela limpeza dos dados segue o fluxo de quatro passos apresentado na Figura 54.

Figura 54 – FLUXOGRAMA DE LIMPEZA DE DADOS FAO



Fonte: o Autor (2018)

A limpeza dos dados se inicia numa iteração para percorrer todos os objetos contidos na página com o intuito de retirar os dados de cada um deles (linhas 1.061 à 1.146 do Apêndice E). Antes da inserção dos dados no *data frame*, o algoritmo

uma realiza a conversão dos dados de unidade (entre as linhas 1.080 e 1.089 para agricultura e 1.110 e 1.135 para pecuária no Apêndice E) e valor seguindo as seguintes regras:

- Caso a unidade esteja em hectares, o valor é transformado em mil hectares;
- Caso a o indicador de produtividade esteja na unidade de hectograma por hectare, o valor é alterado para quilogramas por hectare;
- Caso a unidade esteja em toneladas, é substituído por mil toneladas;
- Unidade em toneladas transformada em mil toneladas;
- Unidade em "1000 Head" transformada em "Mil Cabeças", sem a alteração do valor de origem;
- Unidade em "Head" transformado em "Mil Cabeças";
- Unidades da produtividade em 100mg por animal, hectograma, hectograma por animal e 0.1g por animal transformados em quilogramas por animal;
- Unidade em "1000 No" transformado em "Mil unidades", sem alteração no valor original, somente no nome da unidade de medida.

Por fim, as colunas de estado e fonte são preenchidas por padrão, para este algoritmo, com os valores "Brasil" e "FAO". Após a conversão, os dados são inseridos no *data frame* (linhas 1.091 à 1.098 para agricultura e 1.136 à 1.143 para pecuária, presentes no Apêndice E) até o fim da iteração, resultando em um *data frame* como da Figura 55.

Figura 55 – *DATA FRAME* COM DADOS DA FAO PÓS ITERAÇÃO

	ANO	ESTADO	FONTE	PRODUTO	INDICADOR	UNIDADE	VALOR
27	1969	Brasil	FAO	Apples	Area harvested	Mil Hectares	2.770
28	1970	Brasil	FAO	Apples	Area harvested	Mil Hectares	2.880
29	1971	Brasil	FAO	Apples	Area harvested	Mil Hectares	2.918
30	1972	Brasil	FAO	Apples	Area harvested	Mil Hectares	3.080
31	1973	Brasil	FAO	Apples	Area harvested	Mil Hectares	3.781
32	1974	Brasil	FAO	Apples	Area harvested	Mil Hectares	4.422
33	1975	Brasil	FAO	Apples	Area harvested	Mil Hectares	5.123
34	1976	Brasil	FAO	Apples	Area harvested	Mil Hectares	5.996
35	1977	Brasil	FAO	Apples	Area harvested	Mil Hectares	6.593
36	1978	Brasil	FAO	Apples	Area harvested	Mil Hectares	7.183
37	1979	Brasil	FAO	Apples	Area harvested	Mil Hectares	8.484
38	1980	Brasil	FAO	Apples	Area harvested	Mil Hectares	10.401

Fonte: o Autor (2018)

Por fim, para o passo quatro restou a tradução dos dados, que vem em inglês a partir de sua fonte. Foram traduzidos os nomes dos produtos e o nome dos indicadores entre as linhas 1.151 e 1.409 do Apêndice E. O tratamento de possíveis erros acontece na linha 1.413.

4.3.5 Tratamento geral dos dados

Após a realização da coleta e de um primeiro tratamento prévio dos dados, ainda não há uma uniformidade completa nos dados, o que poderia gerar problemas na utilização da ferramenta. Sendo assim, foram realizadas padronizações nos dados nos seguintes sentidos:

- Remoção de dados com valor "0" ou nulo - para garantir que nenhum dado com valor zero ou nulo esteja entre os dados, a planilha é filtrada para retirar possíveis ocorrências (linhas 1.420 à 1.423 do Apêndice E);
- Estado - pela definição da seção 3.3.3 Tratamento dos dados (p.86) o nome do estado deve vir por extenso, porém, dados da EMBRAPA e CONAB utilizam apenas a sigla ao invés do nome do estado. Dessa forma, foi necessário realizar, via algoritmo, a alteração dos registros com sigla para o nome inteiro do estado (entre as linhas 1.426 e 1.460 do Apêndice E);
- Nome de produto - foi necessário padronizar o nome de produtos pelo fato de haverem diferenças no nome dos produtos disponibilizados pelas fontes, apesar de se tratarem do mesmo, como por exemplo "Algodão" e "ALGODÃO", "Feijão (1ª Safra)" e "FEIJÃO 1ª SAFRA" e assim sucessivamente (linhas 1.463 à 1.524 do Apêndice E);
- Indicador - questão semelhante aos produtos, haviam diferenças nos nomes dos indicadores, como por exemplo "Rendimento médio" e "Produtividade", o que causou a necessidade de uniformizar. Após a uniformização, registros com indicadores diferentes dos apresentados no QUADRO 5 (p. 57) foram removidos (linhas 1.527 e 1.541 do Apêndice E);
- Unidade - uniformização para que os dados fiquem todos com o mesmo tipo de unidade, como por exemplo todos os dados de produtividade em kg/hectare, os de produção em mil toneladas, e de área plantada e colhida em mil hectares (linhas 1.544 à 1.550 do Apêndice E);

Assim como a limpeza dos dados das base, a limpeza geral também contempla um *try catch* para que, ao acontecerem erros, ser gerado um arquivo .TXT que indique

que ocorreu um problema na limpeza dos dados nesta parte (linha 1.554 do Apêndice E).

Após a limpeza geral dos dados, são gerados dois arquivos formato .XLS com os dados de pecuária e agricultura (linhas 1.558 e 1.559 do Apêndice E) e também é realizada a publicação da nova versão da ferramenta no ShinyApps, juntamente com as duas planilhas com os dados atualizados (linha 1.560 do Apêndice E).

4.4 DEFINIÇÕES DA FERRAMENTA

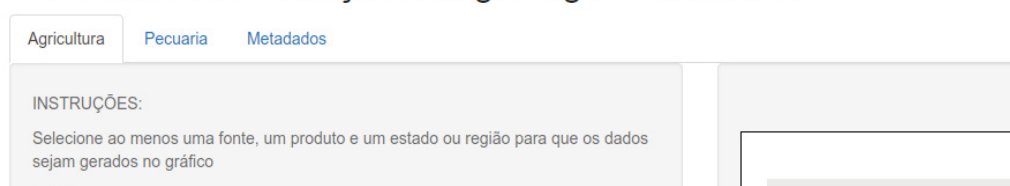
Este capítulo designa-se a apresentar resultados da construção da ferramenta a partir das definições apresentadas na Seção 3.3.4, p. 88. Para o funcionamento da ferramenta descrita nesta seção, utiliza-se os dados coletados e limpos através dos passos descritos nas seções anteriores (Coleta de Dados e Tratamento dos Dados).

A ferramenta conta com uma separação (Figura 56) para a visualização dos dados em três abas (linhas 225 à 229 do Apêndice F) da seguinte forma:

- Agricultura - contempla os filtros específicos da agricultura, bem como quatro quadros com os gráficos (produção, produtividade, área plantada e área colhida) a serem desenhados a partir da seleção dos filtros, bem como quatro quadros para os dados estatísticos da seleção. A função que monta a aba de agricultura encontra-se entre as linhas 17 e 91 do apêndice F;
- Pecuária - assim como na parte da agricultura, também contempla quadros para plotagem dos dados nos gráficos e para indicação de informações estatísticas. São quadros três quadros de cara para os indicadores de rebanho, produção e produtividade. A função que apresenta a aba de pecuária apresenta-se entre as linhas 93 e 156 do Apêndice F;
- Metadados - consta nesta aba os metadados da ferramenta, informados de acordo com o QUADRO 11, p. 90 (linhas 158 à 220 do Apêndice F).

Figura 56 – SEPARAÇÃO DE ABAS DA FERRAMENTA

Indicadores de Produção do Agronegócio Brasileiro



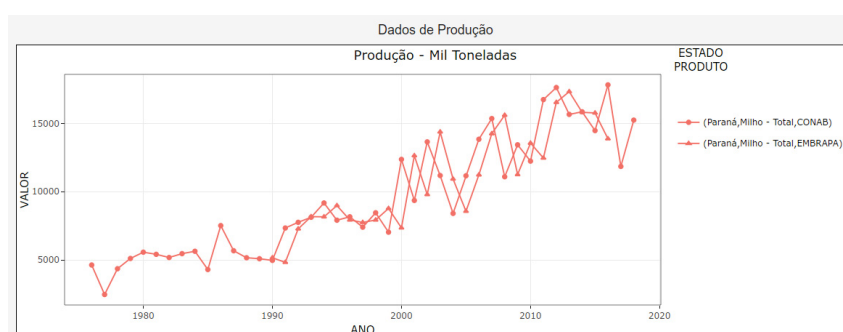
Fonte: o Autor (2019)

4.4.1 Gráficos de Linha

Quanto aos gráficos, a disponibilização dos mesmos é feita através da biblioteca Plotly, com a função ggplotly, conforme indicado pela seção de definições da ferramenta (3.3.4, p. 88). Após cada interação do usuário com os filtros, os dados são filtrados de acordo com as escolhas, e no caso haverem dados com as características selecionadas, é apresentado o gráfico com os dados.

Para o módulo de agricultura, são construídos quatro gráficos, sendo um para cada indicador, conforme a Figura 57, que mostra um exemplo de gráfico de produção de milho gerado para o estado do Paraná utilizando dados da CONAB e EMBRAPA. Para os dados de pecuária, são disponibilizados três gráficos de linha: rebanho, produção e produtividade.

Figura 57 – EXEMPLO DE GRÁFICO GERADO PELA FERRAMENTA



Fonte: o Autor (2019)

A geração de todos os sete gráficos de linha de agricultura e pecuária encontram-se todos disponibilizados no Apêndice F, da seguinte forma:

- Produção - Agricultura - linhas 251 à 267, e a invocação do gráfico na linha 42;
- Produtividade - Agricultura - linhas 304 à 321, e a invocação do gráfico na linha 55;
- Área Plantada - Agricultura - linhas 358 à 373, e a invocação do gráfico na linha 68;
- Área Colhida - Agricultura - linhas 410 à 424, e a invocação do gráfico na linha 81;
- Rebanho - Pecuária - linhas 472 à 488, e a invocação do gráfico na linha 120;
- Produção - Pecuária - linhas 525 à 541, e a invocação do gráfico na linha 133;
- Produtividade - Pecuária - linhas 578 à 595, e a invocação do gráfico na linha 146.

Conforme indicado na seção de definições da ferramenta dentro da metodologia (3.3.4, p. 88), foram construídos sete gráficos pelo fato de terem sido encontrados sete diferentes valores apontados como indicadores: produção, produtividade, área plantada e área colhida para agricultura, e rebanho, produção e produtividade para pecuária.

4.4.2 Métodos Estatísticos e Boxplot

Assim como nos gráficos de linha, os métodos estatísticos são mostrados caso as características selecionadas pelo usuário através dos filtros traga resultados para serem calculados; também são apresentados quatro quadros estatísticos para a pecuária, e três para pecuária.

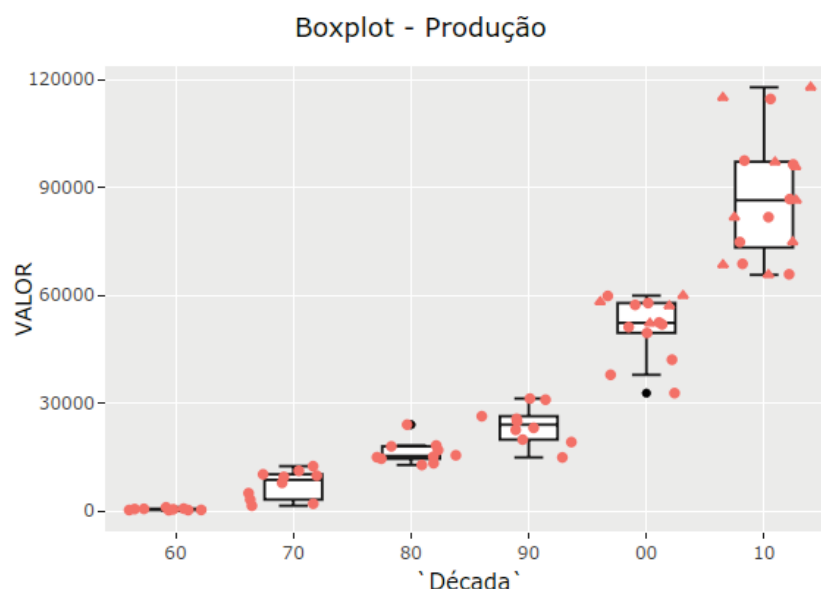
Os dados são apresentados dentro de quadros, logo abaixo dos gráficos de linha, onde são condensando por década conforme definido dentro da metodologia, com intuito de facilitar a visualização da progressão dos dados ao usuário. A Figura 58 mostra o quadro contendo os dados da estatística descritiva, definidos previamente no capítulo da metodologia, de pecuária, sendo resultado da pesquisa do tamanho do rebanho de bois do Mato Grosso.

Figura 58 – EXEMPLO DE TABELA ESTATÍSTICA GERADA PELA FERRAMENTA

Década	Media	Desvio Padrao	Minimo	Maximo
70	3681.61	668.55	2934.17	4222.53
80	6750.32	1028.01	5249.32	8473.93
90	13128.17	4259.09	9041.26	17242.94
00	25963.27	1372.22	24613.72	27357.09
10	29227.86	767.67	28592.18	30296.10

Fonte: o Autor (2019)

Assim como cada os dados estatísticos, os *boxplots*, também são construídos para cada década. A Figura 59 demonstra a construção dos *boxplots* para dados de soja do Brasil segundo a FAO e a EMBRAPA.

Figura 59 – EXEMPLO DE *BOXPLOT* GERADO PELA FERRAMENTA

Fonte: o Autor (2019)

A construção tanto da tabela quanto do *boxplot* se dá através dos seguintes trechos de código apresentados no Apêndice F:

- Produção - Agricultura - linhas 270 à 296;
- Produtividade - Agricultura - linhas 324 à 350;
- Área Plantada - Agricultura - linhas 376 à 402;
- Área Colhida - Agricultura - linhas 427 à 453;
- Rebanho - Pecuária - linhas 491 à 518;
- Produção - Pecuária - linhas 554 à 570;
- Produtividade - Pecuária - linhas 598 à 625.

4.4.3 Filtros

Assim como indicado na Seção 3.3.4, p. 88, a ferramenta disponibiliza ao usuário campos para seleção de acordo com os critérios desejados pelo mesmo para visualização dos dados, são eles: ano, fonte, produto e estado (Figura 60).

Figura 60 – FILTROS DISPONIBILIZADOS PELA FERRAMENTA

Agricultura Pecuária Metadados

INSTRUÇÕES:
Selecione ao menos uma fonte, um produto e um estado ou região para que os dados sejam gerados no gráfico

Anos
1961 2019

Fontes:
☒ IBGE
☒ CONAB
☐ EMBRAPA
☐ FAO

Produtos

Estados

- Acre
- Alagoas
- Amapá
- Amazonas
- Bahia
- Brasil
- Ceará
- Distrito Federal

Fonte: o Autor (2019)

Com relação ao filtro dos anos, o usuário pode selecionar um período temporal para visualização dos dados entre o menor e o maior ano dentro os registros de agricultura e pecuária. A construção dos filtros se encontra entre as linhas 23 e 27 para a aba de agricultura e 99 e 103 da aba de pecuária, ambas presentes no Apêndice F.

Quanto aos filtros de fonte dos dados, são disponibilizados todos os diferentes tipos de fonte encontrados nas planilhas com os dados (linhas 28 à 30 para agricultura e 104 à 106 para pecuária, Apêndice F).

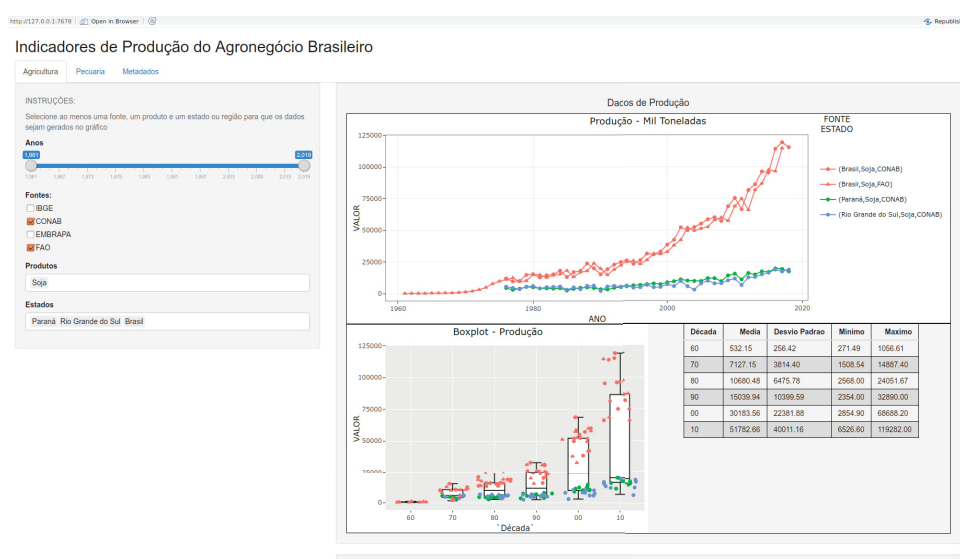
Os produtos e estados são disponibilizados listagens onde o usuário pode também filtrar resultados ao escrever um nome de produto e ir selecionando-os para a consulta que deseja (linhas 31 e 32 para aba de agricultura e 107 à 110 para pecuária, Apêndice F).

Assim que o usuário interage com quaisquer dos filtros disponibilizados, a função nativa do Shiny "reactive" (linhas 235 à 243 para agricultura e linhas 456 à 464 para pecuária, Apêndice F) é acionada, dando início a filtragem dos dados a partir das escolhas do usuário e, no caso de serem encontrados dados com as características desejadas, são construídos os gráficos e informações estatísticas com os dados selecionados.

4.4.4 Dashboard

A ferramenta foi construída utilizando o conceito de *dashboard* conforme indicado na Seção de Definições da Ferramenta (p. 88). A ferramenta (Figura 61) é composta por três abas, que podem ser visualizadas na parte superior. As abas de pecuária e agricultura são constituídas de uma seção na parte esquerda da ferramenta onde são dispostos os filtros para iteração do usuário. Ao lado dos filtros, são apresentadas as informações pertencentes aos filtros selecionados pelo usuário.

Figura 61 – VISÃO GERAL DA FERRAMENTA



Fonte: o Autor (2019)

Quanto aos metadados da ferramenta, a última aba apresenta uma tabela em que constam as informações de: título, assunto, relações, fonte, cobertura, tipo, descrição, criador, colaborador, data, formato e idioma (Figura 62).

Figura 62 – DISPONIBILIZAÇÃO DOS METADADOS DA FERRAMENTA

http://127.0.0.1:7678 | Open in Browser

Indicadores de Produção do Agronegócio Brasileiro

Agricultura Pecuária **Metadados**

Elemento	Valor
Título	Indicadores de Produção do Agronegócio Brasileiro
Assunto	Agronegócio, Indicadores de Produção, Visualização da Informação
Relações	CONAB - Portal de Informações Agropecuárias Observatório Agrícola EMBRAPA - AGROPENSA
Fonte	CONAB, EMBRAPA, FAO, IBGE
Cobertura	1000047
Tipo	Dados relacionados a produção do agronegócio brasileiro
Descrição	A ferramenta tem como objetivo disponibilizar dados do agronegócio brasileiro através de gráficos e filtros interativos
Criador	Jhonny Ivair de Lima Maciel
Colaborador	Prof. Dr. Egon Walter Wildauer
Data	20/02/2019
Formato	Ferramenta web em R utilizando a biblioteca Shiny
Idioma	pt

Fonte: o Autor (2019)

4.4.5 Rotina de captação, tratamento dos dados e publicação da ferramenta

Conforme definido na Seção 3.3.4 Definições da ferramenta (p.88), foi instaurada uma rotina no Cron para que os dados sejam atualizados todos os dias às 00:00, bem como a publicação da ferramenta com os dados atuais. A Figura 63 mostra o trecho que realiza a rotina diária para coleta e limpeza dos dados.

Figura 63 – ROTINA PARA COLETA E LIMPEZA DOS DADOS

```
#
# For more information see the manual pages of crontab(5) and cron(8)
#
# m h dom mon dow  command
0 0 * * * Rscript /home/jhonny/agro/coleta_limpeza_final.R
```

Fonte: o Autor (2018)

4.5 HOSPEDAGEM

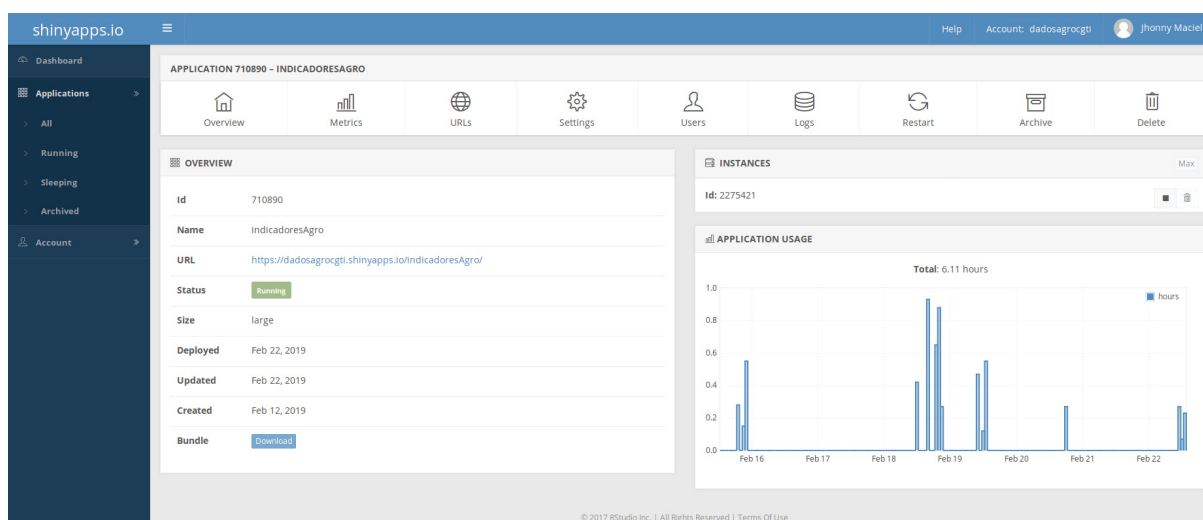
Conforme indicado na Seção 3.3.5, p. 91, a hospedagem da ferramenta no ShinyApps pode ser realizada de duas formas: via interface ou código em R. O presente trabalho efetua a publicação da ferramenta no ShinyApps via código, que é executado diariamente pela rotina do Cron, apresentada na Figura 63.

A autenticação para publicação da ferramenta no ShinyApps se faz presente nas linhas 43 à 45 do Apêndice E, enquanto a publicação da ferramenta ocorre após

todos os procedimentos de coleta e limpeza dos dados (linha 1.561 do apêndice E). Os arquivos submetidos para publicação do ShinyApps são: *script* da ferramenta (Apêndice F) e as planilhas com dados de agricultura e pecuária.

Após a publicação da ferramenta, fica disponível a visualização de estatísticas relacionadas bem como outras ações dentro do ShinyApps (Figura 64). A visualização da ferramenta publicada no ShinyApps pode ser visualizada através do endereço <<https://dadosagrocgti.shinyapps.io/indicadoresAgro/>>.

Figura 64 – ROTINA PARA COLETA E LIMPEZA DOS DADOS



Fonte: o Autor (2019)

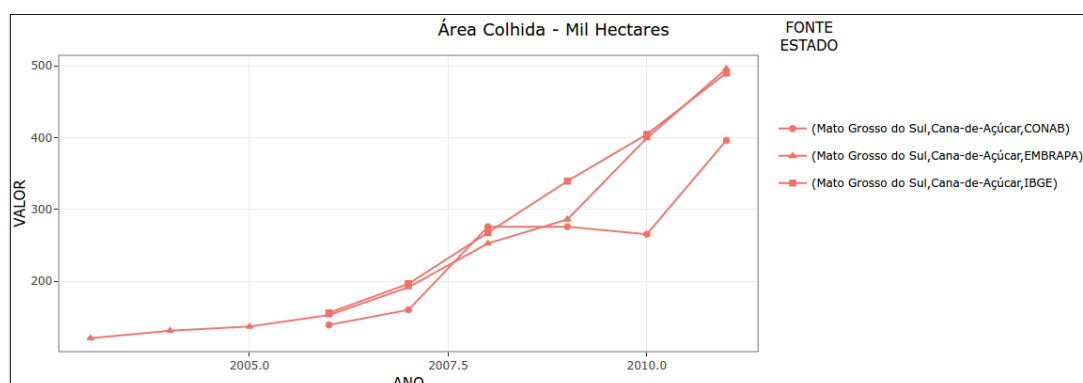
4.6 VALIDAÇÃO DOS DADOS DA FERRAMENTA

A validação dos dados da ferramenta, de acordo com a Seção 3.3.6 (p. 95), ocorreu através de uma espécie de *benchmarking*, comparando dados apresentados na literatura com dados gerados pela ferramenta. Foram selecionados sete trabalhos para a comparação.

- 1 - Caracterização da produção de cana-de-açúcar através de indicadores espaciais e temporais em Mato Grosso do Sul - (MAIA et al., 2015)

O trabalho de Maia et al. (2015) faz referência logo em seu resumo sobre a expansão da área colhida de cana-de-açúcar no estado de Mato Grosso do Sul entre 2003 e 2011. A ferramenta, de acordo com dados do IBGE, EMBRAPA e CONAB, também aponta este crescimento (Figura 65).

Figura 65 – AUMENTO DA ÁREA COLHIDA DE CANA-DE-AÇÚCAR EM MATO GROSSO DO SUL ENTRE 2003 E 2011

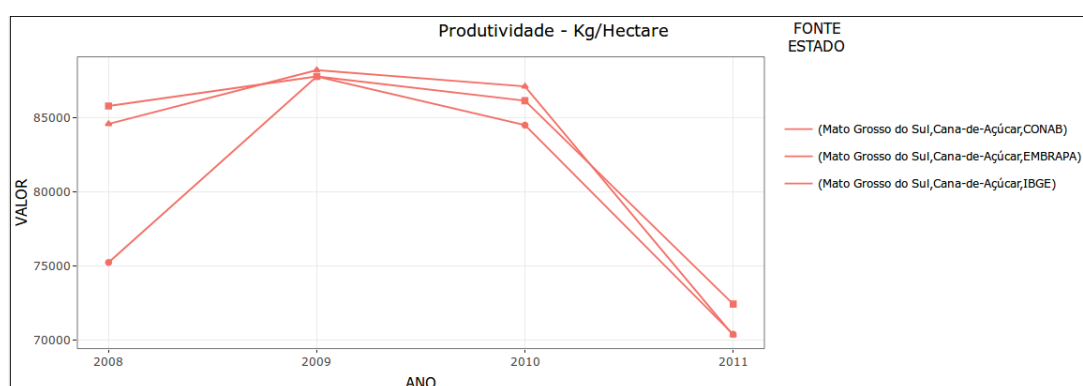


Fonte: o Autor (2019)

Outro dado apontado pelo trabalho, na página 33, é que a área colhida no ano de 2011 foi superior a do ano anterior, 2010, em 85 mil hectares. A Figura 65, logo acima, mostra que a ferramenta também corrobora com o apontamento da pesquisa, indicando o aumento de cerca de 400 mil hectares colhidos para perto de 500 mil de um ano para o outro, de acordo com IBGE e EMBRAPA. Já os dados da CONAB apontam crescimento superior, passando de 265 mil hectares para 396 mil.

Por fim, Maia et al. (2015) citam que a produtividade do estado caiu de 84,5 toneladas por hectare para 70,6 de 2010 para 2011. Os dados da EMBRAPA, IBGE E CONAB mostram a mesma afirmação que os autores (Figura 66).

Figura 66 – VARIAÇÃO DE PRODUTIVIDADE DE CANA-DE-AÇÚCAR ENTRE 2010 E 2011 EM MATO GROSSO DO SUL



Fonte: o Autor (2019)

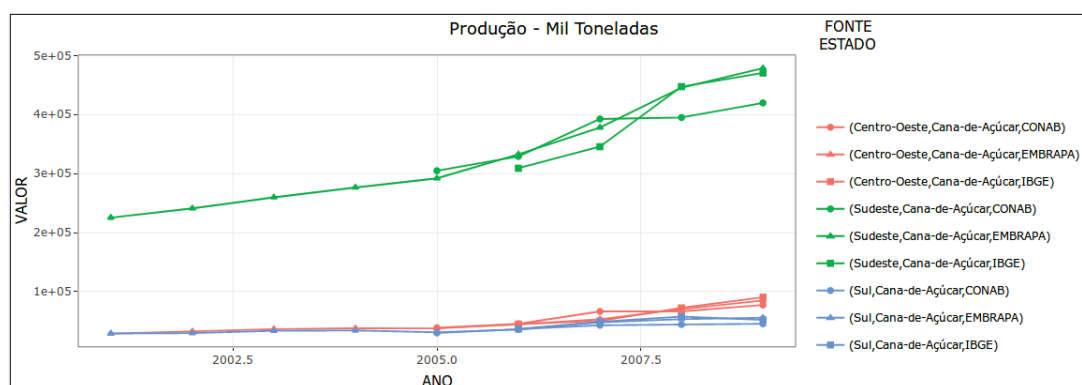
- 2 - Expansão Canavieira e Ocupação Formal em Empresas Sucroalcooleiras do Centro-Sul do Brasil, entre 2007 e 2009 - (BACCARIN; GEBARA; BORGES

JUNIOR, 2011)

Os autores apontam em seu trabalho (p. 494) que a produção de cana-de-açúcar no centro-sul do Brasil (centro-oeste, sudeste e sul) passou de 205,2 milhões de toneladas em 2001 para 500,2 milhões de toneladas em 2009 de acordo com dados do Ministério de Agricultura, Pecuária e Abastecimento (MAPA). A ferramenta aponta (Figura 67), através de dados da EMBRAPA um total de 283,4 milhões de toneladas nas três regiões (225,4 milhões no sudeste, 28,9 milhões no sul e 29,1 milhões no centro-oeste), resultado cerca de 41% maior do que o apresentado pelos autores no trabalho. Já os dados de 2009 são mais similares:

- EMBRAPA - 619,3 milhões de toneladas produzidas (478,5 milhões no sudeste, 85,1 milhões no centro-oeste e 55,7 milhões no sul), valor 23% maior que o apresentado na pesquisa;
- CONAB - 504,8 milhões de toneladas (395 milhões de toneladas no sudeste, 65,5 milhões no centro-oeste e 44,3 milhões no sul), valor 0,91% maior que o do MAPA;
- IBGE - 613,4 milhões de toneladas (470,6 milhões de toneladas no sudeste, 90,7 milhões no centro-oeste e 52,1 no sul), valor 22% maior do que o indicado pelos autores na pesquisa;

Figura 67 – VARIAÇÃO DE PRODUÇÃO DE CANA-DE-AÇÚCAR ENTRE 2001 E 2009 NO CENTRO-SUL DO BRASIL



Fonte: o Autor (2019)

- 3 - Evolução recente e tendências do agronegócio - (CONTINI et al., 2006)

O trabalho de Contini et al. (2006) cita algumas comparações de indicadores de grãos entre Brasil, Argentina e Estados Unidos apresentando números. Uma das

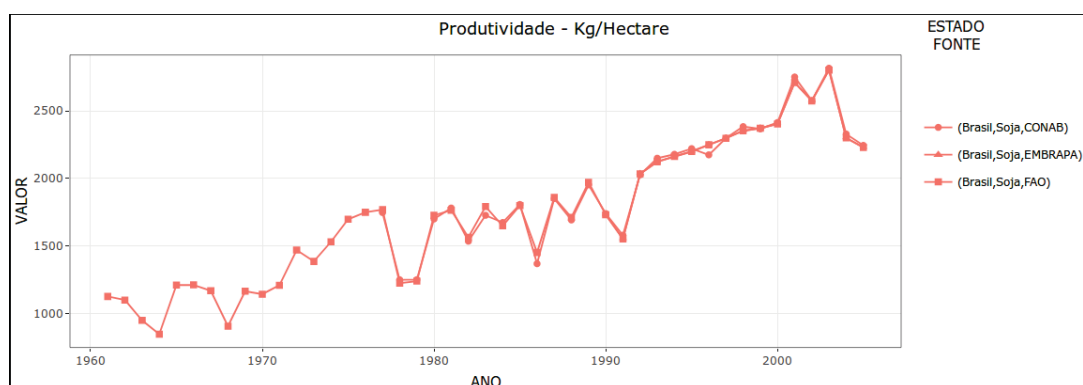
comparações, na página 8, apresenta a evolução da produtividade da soja entre os três países nos anos de 1964, 1974, 1984, 1994, 2004 e 2005 utilizando dados do Departamento de Agricultura dos Estados Unidos (USDA). A comparação dos resultados obtidos pela ferramenta (Figura 68) e pelo trabalho dos autores é apresentado na Tabela 5.

Tabela 5 – COMPARAÇÃO DE RESULTADOS DA FERRAMENTA E DO TRABALHO DE CONTINI ET AL., 2006 - PRODUTIVIDADE DE SOJA EM TONELADAS POR HECTARE NO BRASIL

Ano	Contini et al. (2006)	FAO	CONAB	EMBRAPA	IBGE
1964	1,2	0,8	-	-	-
1974	1,7	1,5	-	-	-
1984	1,8	1,6	1,6	-	-
1994	2,2	2,2	2,1	2,3	-
2004	2,4	2,3	2,3	2,3	-
2005	2,7	2,2	2,2	2,2	-

Fonte: o Autor (2019)

Figura 68 – RESULTADOS DA FERRAMENTA PARA PRODUTIVIDADE DE SOJA NO BRASIL ENTRE 1964 E 2005



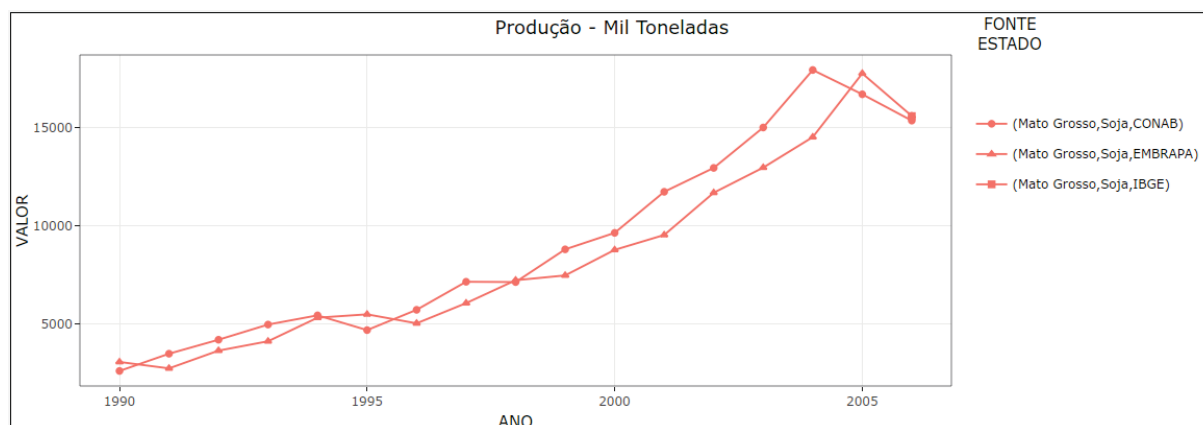
Fonte: o Autor (2019)

- 4 - Análise da competitividade do algodão mato-grossense no cenário internacional: 1999 a 2005 - (SOUSA; BONJOUR; FIGUEIREDO, 2006)

Os autores afirmam, na página 190, que houve aumento considerável, embora não afirme de quanto é o "aumento considerável", na produção de soja no estado de Mato Grosso a partir de 1997 decorrente de criação da Lei Kandir de Setembro de

1996. De acordo com os resultados da ferramenta (Figura 69) realmente houve um aumento conforme indicado pelos autores.

Figura 69 – AUMENTO NA PRODUÇÃO DE SOJA NO ESTADO DE MATO GROSSO A PARTIR DO ANO DE 1997



Fonte: o Autor (2019)

O trabalho também aponta, apresentando o IBGE como fonte, uma tabela (p.193) com dados específicos de área colhida, produção e produtividade de soja de 1990 a 2006 tanto do estado de Mato Grosso como do Brasil. Para a validação dos dados da ferramenta, foram cruzados estes dados divulgados na pesquisa com os dados apresentados pela ferramenta. Os resultados se encontram no Apêndice G. Foram realizadas as seguintes comparações:

- Área Colhida de soja no estado do Mato Grosso entre 1990 e 2006 (Tabela 11);
- Produção de soja no estado do Mato Grosso entre 1990 e 2006 (Tabela 12);
- Produtividade de soja no estado do Mato Grosso entre 1990 e 2006 (Tabela 13);

No que diz respeito aos dados de soja do estado do Mato Grosso, os resultados apontados pela ferramenta com dados do IBGE e CONAB foram similares aos apresentados pela pesquisa de Sousa, Bonjour e Figueiredo (2006). Os dados que apresentaram variação maior foram os da CONAB, mas, apesar da variação, mostraram um comportamento semelhante.

Os dados de produção do trabalho dos autores apontam o crescimento de produção de cerca de três milhões de toneladas para 15,5 milhões entre 1990 e 2006. Os dados da CONAB, apesar das variações apresentadas na Tabela 12, demonstra um crescimento parecido: 2,9 milhões de toneladas para 16,7 milhões no mesmo período da pesquisa. Quanto a produtividade, ambas apontam crescimento: de duas toneladas

por hectare para aproximadamente 2,7 de acordo com a pesquisa de Sousa, Bonjour e Figueiredo (2006), e, 1,9 toneladas por hectare para aproximadamente 2,7 de acordo com a ferramenta com dados da CONAB.

As comparações de área colhida, produção e produtividade também foram realizadas (Tabela 14, Tabela 15 e Tabela 16 respectivamente) a nível Brasil; os resultados das comparações estão disponíveis no Apêndice G. A comparação dos dados de área colhida do trabalho de (SOUSA; BONJOUR; FIGUEIREDO, 2006) e da ferramenta através de dados do CONAB, IBGE, EMBRAPA E FAO. Apenas no ano de 2000 todas as fontes indicam variações superiores a 15%, nos demais dados, a maior variação encontrada foi de 5,1%.

Com relação a produção e a produtividade de soja no Brasil, os dados apresentados no trabalho mais uma vez mostram-se semelhantes aos dados da ferramenta obtidos do CONAB, IBGE, EMBRAPA e FAO, com variações inferiores a 3,5%.

A pesquisa também apresenta dados relacionados a área colhida, produção e produtividade de algodão em caroço no estado de Mato Grosso e no Brasil. Das quatro fontes de dados do presente trabalho, somente a CONAB disponibiliza resultados de algodão em caroço de produção e produtividade. A Tabela 17 e a Tabela 18, presentes no Apêndice H, mostram a comparação dos dados apresentados por Sousa, Bonjour e Figueiredo (2006) e os dados da CONAB apresentados pela ferramenta.

Apesar da discrepância apresentada em alguns valores na comparação dos dados, ambos apontam crescimento da produção de algodão em caroço no Mato Grosso entre 1990 e 2006: aumento de cerca de 24,6 vezes segundo os autores do trabalho e de 22,6 vezes segundo os dados da CONAB utilizados na ferramenta. Com relação a produtividade, também apresentaram-se variações nos resultados, porém, ambos afirmam que o crescimento no período girou em torno de 2,3 toneladas por hectare entre 1990 e 2006.

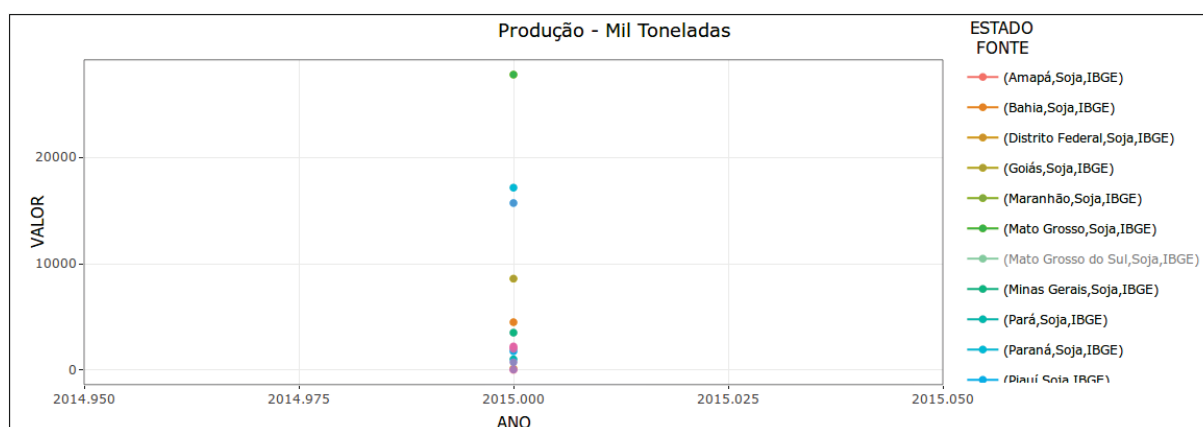
Da mesma forma que os dados de algodão em caroço do estado do Mato Grosso, a Tabela 19 e a Tabela 20 do Apêndice H relatam que também foram encontradas variações nos resultados de produção e produtividade de algodão em caroço a nível nacional. Igualmente ao caso anterior, apesar das variações que chegaram a ultrapassar os 20% em alguns casos, ambas as fontes convergem em apontar o aumento dos valores encontrados para ambos os indicadores.

- 5 - Agronegócio, produção de alimentos e segurança alimentar na América Latina - (COSTA, 2016)

O trabalho aborda questões relacionadas à produção de soja tanto no Brasil quanto no exterior. O autor afirma, na página 153, que em 2015, 17 estados brasileiros

produziram soja. Ao solicitar na ferramenta dados de soja nos ano de 2015 para todos os estados brasileiros e a fonte IBGE, a ferramenta aponta 17 resultados, assim como indicado na pesquisa de Costa (2016) (Figura 70).

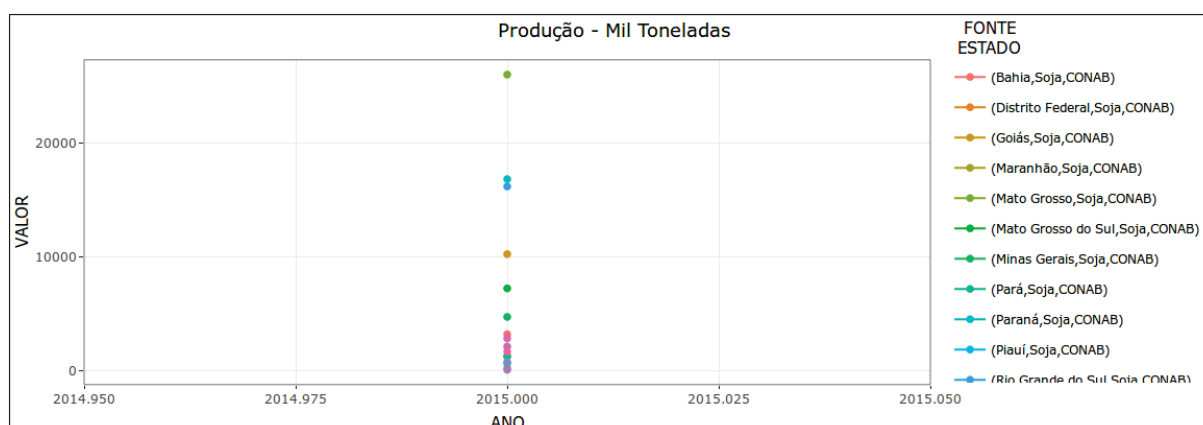
Figura 70 – ESTADOS PRODUTORES DE SOJA EM 2015 SEGUNDO A FERRAMENTA ATRAVÉS DE DADOS DO IBGE



Fonte: o Autor (2019)

Ao selecionar os mesmos filtros, alterando apenas a fonte para CONAB, a ferramenta aponta 16 estados resultantes (Figura 71). A diferença na contagem entre a CONAB e o IBGE está na representação do estado de Amapá, presente apenas nos dados do IBGE.

Figura 71 – ESTADOS PRODUTORES DE SOJA EM 2015 SEGUNDO A FERRAMENTA ATRAVÉS DE DADOS DA CONAB

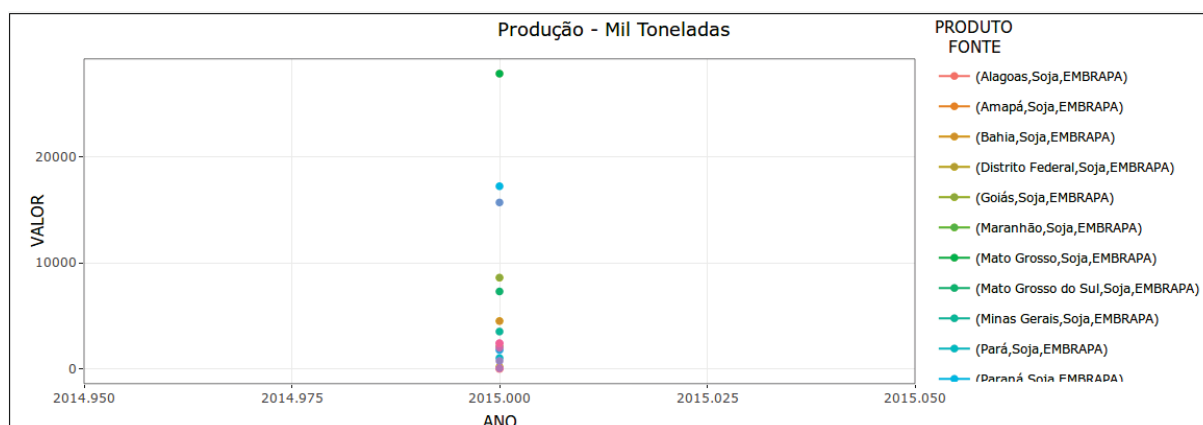


Fonte: o Autor (2019)

Por fim, ao selecionar a EMBRAPA como fonte, são disponibilizados dados de 18 estados diferentes para soja (Figura 72). O estado listado nos dados da EMBRAPA

que não figura para outras fontes é Alagoas.

Figura 72 – ESTADOS PRODUTORES DE SOJA EM 2015 SEGUNDO A EMBRAPA UTILIZANDO A FERRAMENTA



Fonte: o Autor (2019)

Ainda na página 153, o autor afirma que de toda a soja produzida pelo Brasil, o centro-oeste foi responsável por 45,7%, 35% o sul, 8,4% nordeste, 6,1% sudeste e 4,45% a região norte. A Tabela 6 mostra os resultados obtidos pela ferramenta.

Tabela 6 – CONTRIBUIÇÃO POR REGIÃO NA PRODUÇÃO DE SOJA DO BRASIL EM 2015

Região	IBGE (% do total)	CONAB (% do total produzido)	EMBRAPA (% do total produzido)
Centro-Oeste	45,14%	45,69%	45,08%
Nordeste	8,63%	8,04%	8,06%
Norte	4,38%	4,45%	4,38%
Sudeste	5,92%	6,10%	6,08%
Sul	35,90%	35,34%	35,83%

Fonte: o Autor (2019)

O autor também aponta em seu trabalho dados da CONAB de área plantada e produção total por regiões do Brasil. A Tabela 21 e a Tabela 22 do Apêndice I mostram a comparação dos dados da ferramenta com os indicados por Costa (2016) e a variação entre eles. Os dados de área plantada mostraram citados na pesquisa e mostrados pela ferramenta mostraram variações sempre acima dos 35%, enquanto os dados de produção mostraram variação menor, sendo 3,7% a maior encontrada.

Os dados de área plantada indicados no trabalho, e que apresentaram índices de variação acima dos 35%, são oriundos da CONAB. Devido à alta taxa de variação nos valores, foi realizada uma conferência nos dados originais disponibilizados pela

CONAB, com o intuito de comprovar que a ferramenta manteve os valores originais disponibilizados pela divulgadora. De acordo com a planilha original com dados da CONAB, obtida em 20 de Fevereiro de 2019, a área plantada de soja no Brasil sempre ficou abaixo de 36 milhões de hectares (Figura 73), o que comprova a fidelidade da ferramenta com os dados originais disponibilizados pela CONAB.

Figura 73 – DADOS DE ÁREA PLANTADA DE SOJA DISPONIBILIZADOS PELA CONAB

6	REGIÃO/UF	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18	2018/19 Previsão (*)
7	NORTE	574,9	645,5	717,6	901,5	1.178,9	1.441,2	1.576,3	1.809,0	1.931,7	1.997,5
8	RR	1,4	3,7	3,7	12,0	18,0	23,8	24,0	30,0	38,2	48,0
9	RO	122,3	132,3	143,5	167,7	191,1	231,5	252,6	296,0	333,6	333,6
10	AC	-	-	-	-	-	-	-	-	0,5	1,0
11	AM	-	-	-	-	-	-	-	-	1,5	1,9
12	AP	-	-	-	-	-	-	-	18,9	20,2	20,2
13	PA	86,9	104,8	119,2	172,2	221,4	336,3	428,9	500,1	549,6	557,3
14	TO	364,3	404,7	451,2	549,6	748,4	849,6	870,8	964,0	988,1	1.035,5
15	NORDESTE	1.861,7	1.945,7	2.117,1	2.414,3	2.602,2	2.845,3	2.878,2	3.095,8	3.263,5	3.301,3
16	MA	502,1	518,2	559,7	586,0	662,2	749,6	786,3	821,7	951,5	970,5
17	PI	343,1	383,6	444,6	546,4	627,3	673,7	565,0	693,8	710,5	758,1
18	CE	-	-	-	-	-	-	-	-	-	-
19	RN	-	-	-	-	-	-	-	-	-	-
20	PB	-	-	-	-	-	-	-	-	-	-
21	PE	-	-	-	-	-	-	-	-	-	-
22	AL	-	-	-	-	-	-	-	-	2,2	2,2
23	SE	-	-	-	-	-	-	-	-	-	-
24	BA	1.016,5	1.043,9	1.112,8	1.281,9	1.312,7	1.422,0	1.526,9	1.580,3	1.599,3	1.570,5
25	CENTRO-OESTE	10.539,2	10.819,4	11.495,2	12.778,2	13.909,4	14.616,1	14.925,1	15.193,6	15.648,8	16.067,1
26	MT	6.224,50	6.398,50	6.980,50	7.810,20	8.615,70	8.934,50	9.140,0	9.322,8	9.518,5	9.699,5
27	MS	1.712,2	1.760,1	1.815,0	2.017,0	2.120,0	2.300,5	2.430,0	2.522,3	2.672,0	2.816,3
28	GO	2.549,5	2.605,6	2.644,7	2.888,0	3.101,7	3.325,0	3.285,1	3.278,5	3.386,7	3.478,1
29	DF	53,00	54,90	55,00	55,00	72,00	56,10	70,00	70,00	71,50	73,20
30	SUDESTE	1.591,2	1.636,9	1.606,2	1.758,2	1.989,9	2.116,2	2.326,9	2.351,4	2.470,1	2.554,1
31	MG	1.019,0	1.024,1	1.024,0	1.121,2	1.238,2	1.319,4	1.469,3	1.456,1	1.508,5	1.528,1
32	ES	-	-	-	-	-	-	-	-	-	-
33	RJ	-	-	-	-	-	-	-	-	-	-
34	SP	572,2	612,8	582,2	637,0	751,7	796,8	857,6	895,3	961,6	1.026,0
35	SUL	8.900,9	9.133,5	9.106,1	9.883,9	10.492,7	11.074,1	11.545,4	11.459,6	11.835,1	11.901,4
36	PR	4.485,1	4.590,5	4.460,6	4.752,8	5.010,4	5.224,8	5.451,3	5.249,6	5.464,8	5.459,3
37	SC	439,6	458,2	448,3	512,5	542,7	600,1	639,1	640,4	678,2	664,6
38	RS	3.976,2	4.084,8	4.197,2	4.618,6	4.939,6	5.249,2	5.455,0	5.569,6	5.692,1	5.777,5
39	NORTE/NORDESTE	2.436,6	2.591,2	2.834,7	3.315,8	3.781,1	4.286,5	4.454,5	4.904,8	5.195,2	5.298,8
40	CENTRO-SUL	21.031,3	21.589,8	22.207,5	24.420,3	26.392,0	27.806,4	28.797,4	29.004,6	29.954,0	30.522,6
41	BRASIL	23.467,9	24.181,0	25.042,2	27.736,1	30.173,1	32.092,9	33.251,9	33.909,4	35.149,2	35.821,4

Fonte: CONAB (2019)

- 6 - O agronegócio no Brasil e a produção capitalista do território - (FERREIRA, 2012)

O trabalho de Ferreira (2012) apresenta, na página 70, dados relativos à produção do agronegócio Brasileiro. Foi realizada a comparação dos dados indicados no trabalho com os que existem na ferramenta no ano de 2006, mesmo ano dos indicados pelo autor (Tabela 7).

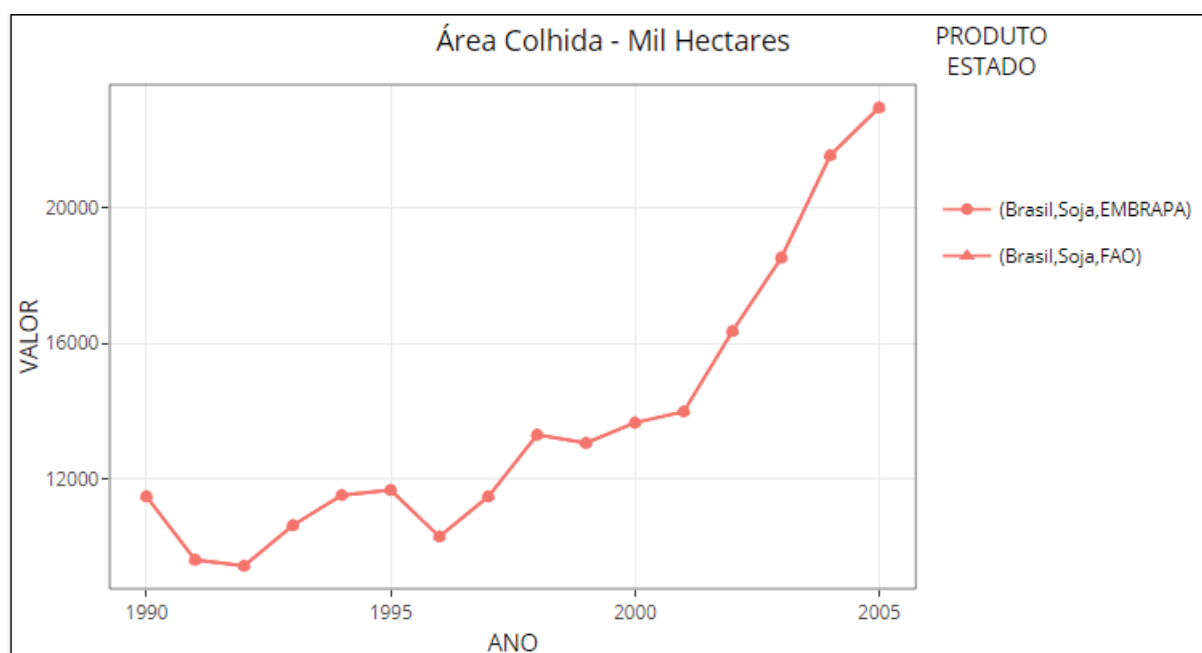
Tabela 7 – COMPARAÇÃO DE RESULTADOS DE PRODUÇÃO INDICADOS NO TRABALHO DE FERREIRA 2012 E DA FERRAMENTA - DADOS DO BRASIL EM 2006

Produto	Contini e Martha Jr. - 2011	IBGE (%)	EMBRAPA (%)	CONAB (%)	FAO (%)
Soja (mil toneladas)	50.000	52.234,59 (4,46%)	52.464,64 (4,92%)	55.027,1 (10,05%)	52.464,64 (4,92%)
Suínos (mil toneladas)	2800	2.298,24 (-17,92%)	-	-	2.830 (1,07%)
Algodão (mil toneladas)	2.700	-	-	-	1.826,6 (-32,34%)
Leite (mil litros)	25.000.000	16.669.743 (-33,32%)	-	-	-
Laranja (mil toneladas)	17.000	18.054,97 (6,2%)	18.032,31 (6,07%)	-	18.032,31 (6,07%)

Fonte: o Autor (2019)

Um gráfico, disposto na página 71 da pesquisa de Ferreira (2012), aponta que a área colhida de soja subiu de cerca de 10 milhões de hectares no ano de 1990 para cerca de 22 em 2005. A ferramenta, através de dados da EMBRAPA e da FAO aponta crescimento similar (Figura 74).

Figura 74 – EVOLUÇÃO DA ÁREA PLANTA DE SOJA NO BRASIL ENTRE 1990 E 2005



Fonte: o Autor (2019)

- 7 - Desempenho e crescimento do agronegócio no Brasil - (GASQUES et al.,

2004)

O trabalho de Gasques et al. (2004) apresenta diversos dados relacionados ao agronegócio brasileiro, sendo grande parte deles de cunho financeiro. Na página 24, porém, é apresentada uma tabela contendo dados relacionados a produção, produtividade e área colhida de: milho, soja, arroz, feijão e trigo na safra de 2002-2003.

A comparação dos dados de produção apontados no trabalho de Gasques et al. (2004) e da ferramenta (Tabela 8) demonstrou uma oscilação entre 1,01% e 1,39% nos resultados de soja. Os dados de milho da ferramenta apontados pela ferramenta mostra uma oscilação maior que a de soja: entre 2,39% e 4,37%.

Os dados de produção de arroz da EMBRAPA, CONAB e FAO adquiridos pela ferramenta mostraram-se ligeiramente maiores que os apresentados pelos autores, sendo sempre menores do que 1%. Os dados de feijão oscilaram entre -12,40% para a FAO e -2,88% para a EMBRAPA. Por fim, os dados de trigo apresentados pela ferramenta mostram valores parecidos entre si, todos distantes cerca de 15% do valor disponibilizados no trabalho de Gasques et al. (2004), apresentando a maior variação entre todos os dados comparados.

Tabela 8 – COMPARAÇÃO DE RESULTADOS DE PRODUÇÃO (EM MIL TONELADAS) INDICADOS NO TRABALHO DE GASQUES ET. AL. 2004 E DA FERRAMENTA - DADOS DO BRASIL EM 2003

Produto	Gasques et al., 2004)	EMBRAPA (%)	CONAB (%)	FAO (%)
Soja	51.300	51.919,44 (1,01%)	52.017,50 (1,39%)	51.919,44 (1,01%)
Milho	46.300	48.327,32 (4,37%)	47.410,90 (2,39%)	48.327,32 (4,37%)
Arroz	10.300	10.334,60 (0,33%)	10.367,10 (0,65%)	10.334,60 (0,33%)
Feijão	3.400	3.302,03 (-2,88%)	-	2.978,3 (-12,40%)
Trigo	5.300	6.073,5 (14,59%)	6.073,5 (14,59%)	6.073,5 (14,59%)

Fonte: o Autor (2019)

A comparação dos dados de área colhida se fez com informações advindas da EMBRAPA e FAO (Tabela 9). Os dados relacionados a soja, milho e arroz mostraram oscilações que não ultrapassaram os 2,5%. Já os resultados da safra de feijão e trigo apresentaram discrepância de 16,87% e 11,31% respectivamente.

Tabela 9 – COMPARAÇÃO DE RESULTADOS DE ÁREA COLHIDA (EM MIL HECTARES) INDICADOS NO TRABALHO DE GASQUES ET. AL. 2004 E DA FERRAMENTA - DADOS DO BRASIL EM 2003

Produto	Gasques et al., 2004)	EMBRAPA (%)	FAO (%)
Soja	18.100	18.524,77 (2,35%)	18524,77 (2.35%)
Milho	12.700	12.956,68 (2,02%)	12.956,68 (2,02%)
Arroz	3.200	3.180,859 (-0,6%)	3.180,859 (-0,6%)
Feijão	3.500	4.090,568 (16,87%)	-
Trigo	2.300	2560,23 (11,31%)	2560,23 (11,31%)

Fonte: o Autor (2019)

Por fim, a comparação dos dados de produtividade entre a ferramenta e o trabalho de Gasques et al. (2004) (Tabela 10) demonstrou variação nos dados de soja inferiores a 1%. Já a variação dos resultados de milho oscilaram entre 5,44% e 9,63%.

Os dados de arroz da EMBRAPA, CONAB e da FAO se mostraram, em geral, próximos ao valor 1,5% menores do que o do trabalho de Gasques et al. (2004). A produtividade do feijão apresentou variação de 15,29% e 4,71% para EMBRAPA e CONAB respectivamente. Os registros de trigo apresentaram as maiores variações de comparações de produtividade com dados da EMBRAPA e FAO: 20,17% e 20,18% respectivamente e 11,35% para CONAB.

Tabela 10 – COMPARAÇÃO DE RESULTADOS DE PRODUTIVIDADE (KG/HA) INDICADOS NO TRABALHO DE GASQUES ET. AL. 2004 E DA FERRAMENTA - DADOS DO BRASIL EM 2003

Produto	Gasques et al., 2004)	EMBRAPA (%)	CONAB (%)	FAO (%)
Soja	2.800	2.802,7 (0,1%)	2.816 (0,57%)	2.802,7 (0,10%)
Milho	3.400	3.727,32 (9,63%)	3.585 (5,44%)	3.727,3 (9,63%)
Arroz	3.300	3.249 (-1,55%)	3.254 (-1,39%)	3.249 (-1,55%)
Feijão	700	807 (15,29%)	732 (4,71%)	-
Trigo	2.000	2.403,494 (20,17%)	2.227 (11,35%)	2.403,5 (20,18%)

Fonte: o Autor (2019)

5 CONCLUSÃO

Este capítulo contempla as considerações finais da pesquisa com relação a cada um dos objetivos listados na Seção 1.2 (p. 22). Além da conclusão sobre os objetivos, também apresenta-se uma consideração final sobre o trabalho, bem como indicações para trabalhos futuros.

5.1 OBJETIVOS DE PESQUISA

O presente trabalho apresentou a questão: **”como disponibilizar uma ferramenta computacional que colete dados de forma automática a partir de fontes da internet, disponibilizando ao usuário indicadores de desempenho de produção do agronegócio brasileiro?”**; e teve como objetivo principal o desenvolvimento de uma ferramenta computacional que possibilite a visualização e filtragem de dados oriundos de fontes relacionadas ao agronegócio brasileiro.

Para o alcance do objetivo proposto para este estudo, foram propostos seis objetivos específicos, que ao serem alcançados, culminariam no atingimento do foco principal do trabalho.

O primeiro objetivo dizia respeito a identificação de fontes de dados do agronegócio do Brasil. Assim sendo, foram identificados, na Seção 4.1 (p. 96), sete portais com dados do agronegócio, dos quais, quatro desses disponibilizam informações pertinentes para o agronegócio, sendo classificados como fontes da matéria prima utilizada posteriormente pela ferramenta: os dados

O segundo objetivo específico se tratava da proposição de um modelo de extração automática dos dados de cada uma das quatro fontes, resultou em um *script* que obtém os dados de acordo com a especificidade de cada um dos portais:

- CONAB - através de *web scraping* para obtenção do *link* que possibilita o *download* das planilhas com os dados;
- FAO - acesso aos endereços eletrônicos que disponibilizam os dados em formato JSON;
- IBGE - acesso direto aos endereços que fazem a requisição para baixar os arquivos com os dados.

Das quatro fontes listadas no primeiro objetivo, apenas os dados da EMBRAPA não puderam ser obtidos de forma automática, pelo fato de serem disponibilizados em

uma ferramenta construída com a tecnologia Flash, o que acabou forçando a coleta manual dos dados da EMBRAPA neste trabalho.

O cumprimento deste objetivo possibilitou a obtenção dos dados, matéria-prima principal para a construção da ferramenta proposta através do objetivo principal. A Seção 4.2 (p. 102) mostra os resultados da resolução do segundo objetivo específico.

Após a obtenção dos dados de forma automática através de um algoritmo que realizasse esta ação, o terceiro objetivo específico designou-se a realizar uma padronização dos dados que vem, a partir de suas distintas fontes, com diferentes formatos.

Para isso, os dados foram submetidos, também através de um *script*, a uma formatação para que se encaixassem no modelo disponibilizado no QUADRO 10 (pagina 87) e que por consequência pudessem posteriormente serem utilizados de forma igual pela ferramenta independentemente de sua fonte.

O segundo e terceiro objetivos foram unidos em um único algoritmo (Apêndice E), para que, assim que os dados de uma fonte fossem obtidos, já sejam também formatados segundo modelo definido. O algoritmo tem um tempo de execução médio de duas horas, sendo que mais da metade deste tempo destina-se ao *download* das planilhas com dados do IBGE, que necessitam um tempo maior de resposta do que a CONAB por exemplo, que trabalha com um modelo similar de requisição de arquivos com os dados através de um *link*.

A construção do algoritmo que realiza a coleta e limpeza dos dados de pecuária e agricultura por fim resultou em dois arquivos formato .XLSX que contém os dados formatados de todas as fontes de ambas as vertentes.

A coleta, análise e tratamento dos dados de agricultura e pecuária realizada no dia 20 de fevereiro de 2019 totalizou 199.771 registros de agricultura, sendo: 50.767 dados da CONAB, 101.597 da EMBRAPA, 11.300 da FAO e 36.107 do IBGE, de um total de 127 produtos. Os dados de pecuária totalizaram 35.892 registros, sendo 21.308 da EMBRAPA, 5.314 da FAO e 9.270 do IBGE, divididos em 86 produtos. Essa gama de dados foi gerada através do que foi apresentado como resultado na Seção 4.3 (p. 111).

O quarto objetivo específico dizia a respeito de definições da ferramenta, como filtros para utilização do usuário, seleção do tipo dos gráficos e métodos estatísticos a serem utilizados e etc. Os filtros disponibilizados pela ferramenta (período temporal, estados, fontes e produtos) possibilitam ao usuário filtrar os dados da maneira que lhe for conveniente, uma das propostas principais da ferramenta.

Quanto aos gráficos de linha utilizados pela ferramenta, estes possibilitam a sobreposição de informações, o que facilita a visualização dos dados nestes casos,

conforme indicado no referencial teórico. Os métodos estatísticos presentes juntamente com os gráficos também auxiliam no entendimento do comportamento geral dos dados obtidos através dos filtros indicados pelo usuário.

Uma das definições da ferramenta diz respeito as rotinas de coleta, tratamento dos dados e publicação da ferramenta com os dados atualizados. Estas rotinas possibilitam que a ferramenta esteja sempre disponível com os dados mais atualizados possível, pois as ações são executadas diariamente. A disponibilização de todos os resultados referentes ao quarto objetivo específico se encontra na Seção 4.4 (p. 124).

A hospedagem da ferramenta, indicada na Seção 4.5 (p. 130), quinto objetivo específico, se fez no servidor ShinyApps, que possibilita a disponibilização de ferramentas Shiny via navegador de internet. A publicação da ferramenta no servidor ShinyApps tornou a tarefa bem mais simples, tendo em vista que este processo é feito em duas linhas de código, uma de autenticação de conta ShinyApps, e a outra que realiza a publicação do aplicativo no servidor.

Por fim, o sexto objetivo específico, que teve seus resultados indicados na Seção 4.6 (p. 131), tratou da validação dos dados da ferramenta, comparando os resultados apontados por ela com dados apresentados em sete diferentes trabalhos utilizados na construção do referencial teórico.

Apesar de em um dos trabalhos os resultados apontarem variações superiores a 50%, grande parcela dos dados mostrou variações inferiores a 10%. Mesmo com variações, todos os aumentos indicados pelos trabalhos foram apontados também pela ferramenta.

A comparação dos resultados demonstrou que há discrepância nos resultados disponibilizados pelas fontes, sejam elas as utilizadas pela ferramenta ou as indicadas nos trabalhos utilizados para a comparação. Esta discrepância, por vezes, dificulta a análise dos dados, pois como indicado anteriormente algumas vezes as variações chegam a superar os 50%.

O cumprimento dos objetivos específicos tornou possível a realização do objetivo principal: a construção de uma ferramenta de visualização da informação que contempla dados de diversas fontes, sempre atualizados e que permite ao usuário a filtragem dos mesmos de acordo com suas necessidades, tudo isso utilizando conceitos indicados pela visualização da informação como gráficos de linha e *dashboard*.

5.2 CONSIDERAÇÃO FINAL

A visualização da informação em ferramentas computacionais pode trazer diversos benefícios para quem a utiliza. Dentre estes benefícios está a vantagem competitiva

para agentes diretamente ligados ao agronegócio, sobretudo na fase “dentro da porteira”, sendo possível realizar um primeiro vislumbre de oportunidades para projetos futuros, tanto no setor privado quanto para políticas públicas.

Uma primeira pesquisa em portais relacionados ao agronegócio e na literatura pertinente mostrou a existência de ferramentas computacionais de visualização da informação, mas demonstrou também a existência de limitações em suas funcionalidades.

Ferramentas como o Agropensa da EMBRAPA e o Portal de Informações Agropecuárias da CONAB possibilitam a visualização da informação, mas apresentam limitações no cruzamento de dados de diferentes produtos e regiões do Brasil, bem como terem como única fonte de dados os gerados pelas próprias instituições; tendo em vista este cenário, o trabalho foi realizado de modo a construir uma ferramenta computacional que pudesse preencher estas lacunas oriundas de limitações que apresentaram as duas ferramentas indicadas anteriormente.

O alcance dos objetivos propostos se deve em muito a construção do referencial teórico, que contribuiu no entendimento de elementos-chave presentes neste trabalho. A primeira seção do capítulo possibilitou o entendimento do agronegócio como um todo, juntamente com o funcionamento de sua cadeia produtiva.

A seção de Fontes de Dados e Informações (2.4, p. 58) trouxe uma gama de possibilidades de tecnologias a serem utilizadas para a construção da ferramenta, bem como metodologias utilizadas, como *tidy data* e uma listagem de sete possíveis fontes de dados do agronegócio. A seção de Visualização da Informação apontou possíveis soluções para a representação dos dados na ferramenta como gráficos de linha e *dashboad*.

O resultado final da ferramenta e toda sua construção demonstram claramente a importância de temas, técnicas e tecnologias abordadas dentro do referencial teórico, vitais para que a ferramenta cumprisse seu papel conforme planejado logo no início da pesquisa.

Mesmo com o alcance do objetivo principal deste trabalho, que se caracteriza pela construção de uma ferramenta que explore lacunas de outras ferramentas já existentes, há de se indicar que ainda existem melhorias que podem ser incrementadas pela ferramenta atual, visando sempre maior apoio a decisão, seja ela no âmbito público, privado e até mesmo no campo da pesquisa.

5.3 SUGESTÕES PARA TRABALHOS FUTUROS

As indicações futuras dividem-se em três tipos de trabalho: a de melhorias e incrementos da própria ferramenta, aproveitando-se de limitações apresentadas

por este, e a construção de futuras pesquisas a partir de indicações dos dados da ferramenta e a replicação da metodologia apresentada para outros ramos de negócio.

Como possíveis incrementos na ferramenta, aponta-se principalmente no aproveitamento de limitações da ferramenta, como por exemplo:

- Incremento de dados de cunho financeiro, que podem aumentar o poder de auxílio a decisão de decisão;
- Incremento no numero de fontes de dados;
- Disponibilizar os dados de nível regional ou até mesmo municipal;
- Melhorar a performance dos algoritmos de coleta e limpeza dos dados;
- Construção de um modelo que consiga realizar a coleta automática dos dados da EMBRAPA, que tiveram de ser colhidos de forma manual;
- Realizar a validação da ferramenta com atores envolvidos no agronegócio brasileiro, tanto no âmbito público como privado;
- Fazer com que a ferramenta forneça, além dos gráficos e dados estatísticos, *insights* sobre o conjunto de dados, auxiliando ainda mais a tomada de decisão por parte do utilizador da ferramenta.

Quanto a segunda indicação, os dados apontam fenômenos que podem ser pesquisados e explicados por trabalhos futuros, seja por meio de pesquisas documentais, quantitativas ou qualitativas.

Por fim, a última indicação diz respeito a replicação da metodologia utilizada para criação de ferramentas de outros ramos, ou até mesmo de outra fase do agronegócio. O modelo pode ser utilizado para construção de ferramentas computacionais do ramo da indústria e comércio, restando aplicar os ajustes apontados pela literatura pertinente.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABBADE, E. B. O papel do agronegócio brasileiro no seu desenvolvimento econômico. *GEPROS. Gestão da Produção, Operações e Sistemas*, Bauru, v. 9, n. 3, p. 149–158, 2014. Citado 3 vezes nas páginas 19, 70 e 85.
- ABIKO, A. K. Tecnologias apropriadas em construção civil. In: *Tecnologias e materiais alternativos de construção*. Campinas: UNICAMP, 2003. p. 55–64. Citado na página 31.
- ADOBE. *Portable Document Format Reference Manual*. [S.l.]: Adobe Systems Inc, 1997. Citado na página 64.
- ADOBE. *O que é PDF?* 2018. Disponível em: <<https://acrobat.adobe.com/br/pt/acrobat/about-adobe-pdf.html>>. Citado na página 64.
- ALBANO, C. S.; WOHLBERG, J.; GARCIA, V. P. Indicadores de desempenho utilizados em beneficiadores de arroz: um estudo na região da campanha-rs. *GEPROS. Gestão da Produção, Operações e Sistemas*, Bauru, n. 4, p. 19–33, 2014. Citado 3 vezes nas páginas 55, 70 e 85.
- ALTOÉ, A.; SILVA, H. O desenvolvimento histórico das novas tecnologias e seu emprego na educação. In: *Educação e Novas Tecnologias*. Maringá: Eduem, 2005. p. 13–25. Citado na página 31.
- ALVES, A. A. et al. Análise de desempenho econômico da produção orgânica de leite: estudo de caso no distrito federal. *Ciência e Agrotecnologia*, Lavras, v. 33, n. 2, p. 567–573, 2009. Citado 2 vezes nas páginas 56 e 57.
- ANGELONI, M. T. Elementos intervenientes na tomada de decisão. *Ciência da Informação*, Brasília, v. 32, p. 17–22, 2003. Citado na página 45.
- ARAÚJO, M. J. *Fundamentos de agronegócios*. São Paulo: Atlas, 2013. Citado 8 vezes nas páginas 20, 33, 34, 36, 37, 38, 39 e 40.
- AUSTER, E.; CHOO, C. W. Ceos, information, and decision making: Scanning the environment for strategic advantage. *Library Trends*, Illinois, v. 43, p. 206–225, 2004. Citado na página 59.
- BACCARIN, J. G.; GEBARA, J. J.; BORGES JUNIOR, J. C. Expansão canavieira e ocupação formal em empresas sucroalcooleiras do centro-sul do brasil, entre 2007 e 2009. *Revista de Economia e Sociologia Rural*, Piracicaba, v. 49, p. 493–506, 2011. Citado 5 vezes nas páginas 58, 70, 85, 95 e 133.
- BANKER, R. D.; POTTER, G.; SRINIVASAN, D. An empirical investigation of an incentive plan that includes nonfinancial performance measures. *The Accounting Review*, p. 65–92, 2000. Citado na página 27.
- BARBOSA, R. R. Monitoração ambiental: uma visão interdisciplinar. *Revista de Administração*, São Paulo, v. 32, n. 4, p. 42–53, 1997. Citado na página 59.

- BARBOSA, R. R. Uso de fontes de informação para a inteligência competitiva: um estudo da influência do porte das empresas sobre o comportamento informacional. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, p. 91–102, 2006. Citado na página 58.
- BARCAROLI, G. et al. Machine learning and statistical inference: the case of istat survey on ict. *Proceeding of 48th scientific meeting of the Italian Statistical Society SIS*, 2016. Citado na página 105.
- BARCAROLI, G. et al. Use of web scraping and text mining techniques in the istat survey on "information and communication technology in enterprises". In: . [S.l.: s.n.], 2014. Citado na página 105.
- BATALHA, M. O. *Gestão agroindustrial: GEPAI: Grupo de Estudos e Pesquisas Agroindustriais*. São Paulo: Atlas, 2001. v. 2. Citado 2 vezes nas páginas 20 e 37.
- BEELEY, C. *Web Application Development with R using Shiny*. [S.l.]: Packt Publishing, 2013. Citado na página 67.
- BEGG, R. Artificial intelligence techniques in medicine and health care. In: *Encyclopedia of Information Science and Technology*. [S.l.: s.n.], 2005. p. 157–162. Citado na página 48.
- BENTANCOURT, S.; ROCHA, R. Metadados de qualidade e visibilidade na comunicação científica. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, Florianópolis, v. 17, 2012. Citado 2 vezes nas páginas 67 e 68.
- BITITCI, U. S.; CARRIE, A. S.; MCDEVITT, L. Integrated performance measurement systems: a development guide. *International Journal of Operations and Production Management*, v. 17, p. 522–534, 1997. Citado na página 54.
- BITITCI, U. S.; CARRIE, A. S.; TURNER, T. Diagnosing the integrity of your performance measurement system. *Control*, v. 24, p. 9–13, 1998. Citado na página 54.
- BITITCI, U. S.; TURNER, T.; BEGEMANN, C. Dynamics of performance measurement systems. *International Journal of Operations and Production Management*, v. 20, p. 692–704, 2000. Citado na página 54.
- BIVAND, R. S.; PEBESMA, E. J.; GÓMEZ-RUBIO, V. *Nova Iorque*; Londres: Springer, 2008. Citado na página 66.
- BOEHMKE, B. *Data Wrangling with R*. 1. ed. [S.l.]: Springer Publishing Company, Incorporated, 2016. Citado na página 66.
- BOTEON, M. *Mercado de informação digital agroêconômica*. Tese (Doutorado) — Universidade de São Paulo - USP, Piracicaba, 2004. Citado 3 vezes nas páginas 21, 69 e 82.
- BRAMBILLA, J.; CARRO, S.; FELÍCIO, M. Descrição e recuperação de notícias jornalísticas por meio de metadados. *Colloquium Exactarum*, Presidente Prudente, v. 8, n. 1, p. 10–21, 2016. Citado na página 67.

- BRASIL. *Agropecuária puxa o PIB de 2017*. 2017. Disponível em: <<http://www.agricultura.gov.br/noticias/agropecuaria-puxa-o-pib-de-2017>>. Citado na página 19.
- BRUM, A. *Modernização da agricultura: trigo e soja*. Ijuí: Fundação Integração Desenvolvimento Educação Noroeste do Estado (FIDENE), 1985. Citado na página 19.
- BUCHELE, R. B. *Business policy in growing firms: A manual for evaluation*. [S.l.]: Chandler Pub. Co, 1967. Citado na página 54.
- BYTHEWAY, A. *The Information Management Body of Knowledge - IMBOK*. Capetown: Cape Technikon, 2004. Citado 3 vezes nas páginas 42, 43 e 47.
- CALDEIRA, J. *100 Indicadores da gestão*. Lisboa: Actual, 2016. Citado 4 vezes nas páginas 20, 51, 52 e 88.
- CALLADO, A. A. C.; CALLADO, A. L. C. Mensuração de desempenho e agronegócio. In: *Agronegócio*. São Paulo: Atlas, 2011. v. 3, p. 133–149. Citado 3 vezes nas páginas 52, 53 e 55.
- CALLADO, A. A. C.; CALLADO, A. L. C. Sistemas agroindustriais. In: *Agronegócio*. São Paulo: Atlas, 2011. v. 3, p. 1–19. Citado 5 vezes nas páginas 37, 38, 39, 40 e 48.
- CALLADO, A. A. C.; CALLADO, A. L. C.; MACHADO, M. A. V. Indicadores de desempenho operacional e econômico: um estudo exploratório no contexto do agronegócio. *Revista de Negócios*, Blumenau, v. 12, n. 1, p. 3–15, 2007. Citado 3 vezes nas páginas 55, 56 e 70.
- CALLADO, A. A. C.; MORAES FILHO, R. A. Gestão empresarial no agronegócio. In: *Agronegócio*. São Paulo: Atlas, 2011. v. 3, p. 20–29. Citado 2 vezes nas páginas 43 e 49.
- CALLADO, A. L.; SOARES, K. R. Análise da utilização de indicadores de desempenho no contexto das agroindústrias. *Custos e @gronegócio*, Recife, v. 10, n. 2, p. 272–284, 2014. Citado 5 vezes nas páginas 19, 23, 51, 52 e 70.
- CALLADO, A. L. C.; CALLADO, A. A.; ALMEIDA, M. A. A utilização de indicadores gerenciais de desempenho industrial no âmbito de agroindústrias. *Revista Eletrônica Sistemas e Gestão*, Niterói, v. 2, p. 102–118, 2007. Citado 3 vezes nas páginas 52, 55 e 70.
- CALLADO, A. L. C.; CALLADO, A. A. C.; ALMEIDA, M. A. A utilização de indicadores de desempenho não-financeiros em organizações agroindustriais: um estudo exploratório. *Organizações Rurais e Agroindustriais*, v. 10, n. 3, p. 35–48, 2008. Citado na página 22.
- CANUTO, A. Agronegócio: a modernização conservadora que gera exclusão pela produtividade. *Revista Nera*, Presidente Prudente, n. 1, p. 1–12, 2004. Citado na página 36.
- CARD, S. K.; MACKINLAY, J. D.; SHNEIDERMAN, B. Readings in information visualization. In: CARD, S. K.; MACKINLAY, J. D.; SHNEIDERMAN, B. (Ed.). San Francisco: Morgan Kaufmann Publishers Inc., 1999. cap. Using Vision to Think, p. 579–581. Citado 3 vezes nas páginas 72, 83 e 88.

CARMO, V. B.; PONTES, C. C. C. Sistemas de informações gerenciais para programa de qualidade total em pequenas empresas da região de campinas. *Ciência e Informação*, Brasília, v. 28, n. 1, p. 49–58, 1999. Citado na página 58.

CASTELLS, M. *Sociedade em Rede*. São Paulo: Paz e Terra, 2005. v. 8. Citado 2 vezes nas páginas 42 e 48.

CHAFFEY, D.; WOOD, S. *Business Information Management Improving Performance Using Information Systems*. Edimburgo: Prentice Hall, 2004. Citado 9 vezes nas páginas 44, 45, 46, 47, 48, 81, 83, 85 e 88.

CHAVIANO, E. R. et al. Indicadores de sostenibilidad agraria en la granja genética el abra de la empresa pecuaria el tablón. *Revista para la transformación agraria sostenibl*, v. 6, p. 7–15, 2018. Citado 3 vezes nas páginas 56, 57 e 58.

CHECKLAND, P.; HOLWELL, S. Data, capta, information and knowledge. In: *Introducing Information Management: the business approach*. Oxford: Elsevier, 2006. p. 47–55. Citado na página 45.

CHEN, C. *Information Visualization Beyond the Horizon*. Londres: Springer, 2006. Citado na página 71.

CHIAVENATO, I. *Introdução à Teoria Geral da Administração*. Rio de Janeiro: Elsevier, 2004. v. 7. Citado na página 48.

CHOO, C. W. *Gestão de informação para a organização inteligente: a arte de explorar o meio ambiente*. Lisboa: Editorial Caminho, 2003. Citado na página 42.

CHOO, C. W. *The Knowing Organization: How Organizations Use Information to Construct Meaning, Create Knowledge, and Make Decisions*. Nova Iorque: Oxford University Press, 2005. v. 2. Citado na página 44.

COELHO, A. L. d. A. L. et al. Avaliação de desempenho organizacional: uma investigação científica das principais ferramentas gerenciais. In: *XV Congresso Brasileiro de Custos*. Curitiba: [s.n.], 2008. Citado na página 50.

COLLIS, J.; HUSSEY, R. *Business Research : A Practical Guide for Undergraduate and Postgraduate Students*. [S.l.]: Palgrave Macmillan, 2003. Citado 4 vezes nas páginas 75, 76, 77 e 78.

CONTINI, E. et al. Evolução recente e tendências do agronegócio. *Revista de Política Agrícola*, Brasília, p. 5–28, 2006. Citado 8 vezes nas páginas 21, 56, 57, 70, 88, 95, 133 e 134.

CONTINI, E.; MARTHA JR., G. B. Brazilian agriculture, its productivity and change. *Bertebos Conference on “Food security and the futures of farms: 2020 and toward 2050”*. Falkenberg, 2010. Citado 3 vezes nas páginas 34, 35 e 70.

CORREA, F. E. *Representação de comercialização agropecuária através do modelo de data warehouse*. Dissertação (Mestrado) — Universidade de São Paulo - USP, São Paulo, 2010. Citado 3 vezes nas páginas 21, 22 e 23.

CORREA, F. E. et al. Data warehouse for soybeans and corn market on brazil. *EFITA conference*, p. 675–681, 2009. Citado na página 21.

CORREA, R. F.; VIEIRA, J. M. d. L. Representações visuais para recuperação de informação na bdtd-ufpe. *Perspectiva em Ciência da Informação*, v. 18, n. 4, p. 18–34, 2013. Citado na página 71.

COSTA, R. A. G. *Utilização de data warehouses para gerenciar dados de redes de sensores sem fio que monitoram polinizadores*. Dissertação (Mestrado) — Universidade de São Paulo - USP, São Paulo, 2011. Citado na página 23.

COSTA, S. M. G. Agronegócio, produção de alimentos e segurança alimentar na América Latina. *Revista de Políticas Públicas*, São Luís, p. 149–156, 2016. Citado 7 vezes nas páginas 34, 37, 70, 95, 136, 137 e 138.

COUTURE-BEIL, A. *JSON for R*. [S.l.], 2018. Citado na página 66.

CRUVINEL, P. E.; ASSAD, E. D. Estudo do mercado brasileiro de software para o agronegócio: Cenários, prospecção e oportunidades. In: MENDES, C.I.C.; OLIVEIRA, D.R.M.S.; SANTOS, A.R. (org.). *Estudo do Mercado Brasileiro de Software para o Agronegócio*. Embrapa Informática Agropecuária. Campinas: [s.n.], 2011. p. 131–146. Citado na página 21.

DAVENPORT, T. R.; PRUSAK, L. *Information ecology: mastering the information and knowledge environment*. Nova Iorque: Oxford University Press, 1997. Citado 2 vezes nas páginas 41 e 42.

DAVIS, J. H.; GOLDBERG, R. A. *A concept of agribusiness*. Boston: Harvard University, 1957. Citado 3 vezes nas páginas 33, 35 e 36.

DECOSTER, S. R. A.; SUN, V. Si, organizações e estratégia. In: *Fundamentos de Sistemas de Informação*. [S.l.]: Elsevier, pages = 3-15, 2014. Citado na página 43.

DEGENT, R. J. A importância estratégica e o funcionamento do serviço de inteligência empresarial. *Revista de Administração de Empresas*, n. 26, p. 78–83, 1986. Citado na página 58.

DERY, K. F.; SAMSON, D. A. Alignment of information technology and human resources strategies. In: *Encyclopedia of Information Science and Technology*. [S.l.: s.n.], 2005. p. 104–110. Citado na página 48.

DIAS, M. P. *A contribuição da visualização da informação para a ciência da informação*. Dissertação (Mestrado) — PUC, Campinas, 2007. Citado 2 vezes nas páginas 76 e 79.

DIAS, M. P.; CARVALHO, J. O. F. A visualização da informação e a sua contribuição para a ciência da informação. *DataGramaZero - Revista de Ciência da Informação*, Rio de Janeiro, v. 8, n. 5, p. 1–16, 2007. Citado 3 vezes nas páginas 70, 71 e 72.

DIXON, J. R.; JR., A. J. N.; VOLLMANN, T. E. *The new performance challenge, measuring operations for word-class competition*. Homewood: [s.n.], 1990. Citado na página 54.

DRUCKER, P. F. The coming of the new organization. *Harvard Business*, Cambridge, 1988. Citado na página 45.

ECKERSON, W.; WHITE, C. *Evaluating ETL and Data Integration Platforms*. [S.l.: s.n.], 2003. Citado na página 60.

ECKERSON, W. W. *Performance Dashboards: Measuring, Monitoring, and Managing Your Business*. New Jersey: John Wiley Sons, Inc, 2006. Citado na página 80.

EHLERS, E. *O que é agricultura sustentável*. São Paulo: Brasiliense, 2009. Citado 3 vezes nas páginas 32, 33 e 39.

EL-SAPPAGH SHAKER H. ALI ANDHENDAWI, A. M. A.; BASTAWISSY, A. H. E. A proposed model for data warehouse etl processes. *Journal of King Saud University – Computer and Information Sciences*, v. 23, p. 91–104, 2011. Citado 2 vezes nas páginas 61 e 62.

ELESBAO, I. O espaço rural brasileiro em transformação. *Finisterra Revista Portuguesa de Geografia*, Lisboa, XLII, p. 47–65, 2007. Citado 2 vezes nas páginas 34 e 70.

ELIAS, D. Redes agroindustriais e organização dispersa no brasil. *X Coloquio Internacional de Geocrítica*, Barcelona, 2008. Citado na página 35.

ELIAS, D. Globalização, agricultura e urbanização no brasil. *Acta Geográfica*, Boa Vista, p. 13–32, 2013. Citado 3 vezes nas páginas 34, 37 e 70.

ELMASRI, R.; NAVATHE, S. B. *Sistemas de banco de dados*. [S.l.: s.n.], 2005. v. 4. Citado na página 59.

ESTIVALET, L. F. *O uso de ícones na Visualização de Informações*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, Porto Alegre, 2000. Citado na página 73.

FABER, M. *A Importância dos Rios para as Primeiras Civilizações*. [S.l.]: www.historialivre.com, 2011. v. 1. Citado na página 31.

FAGUNDES, J. A. et al. Tableau de bord vs balanced scorecard. *Revista de Contabilidade do Mestrado em Ciências Contábeis da UERJ*, Rio de Janeiro, v. 12, p. 1–16, 2007. Citado na página 55.

FERREIRA, G. H. C. O agronegócio no brasil e a produção capitalista do território. *Geografia em Questão*, Marechal Cândido Rondon, v. 28, n. 1, p. 66–82, 2012. Citado 4 vezes nas páginas 36, 95, 139 e 140.

FERREIRA, J. et al. O processo etl em sistemas data warehouse. *INForum 2010 - II Simpósio de Informática*, p. 757–765, 2010. Citado na página 60.

FERREIRA, M. P. et al. Gestão por indicadores de desempenho: resultados na incubadora empresarial tecnológica. *Produção*, v. 18, p. 302 – 318, 2008. Citado na página 20.

FERREIRA, R. d. S.; CAMARGO, S. d. S. Construindo um data warehouse para o agronegócio. *SBIAgro 2013*, Cuiabá, 2013. Citado 5 vezes nas páginas 19, 22, 34, 36 e 70.

FEW, S. *Information Dashboard Design*. Boston: O'Reilly, 2006. Citado 2 vezes nas páginas 79 e 80.

FIGUEIREDO, A. M.; SANTOS, M. L. d.; LIMA, J. F. d. Importância do agronegócio para o crescimento econômico de brasil e estados unidos. *Gestão e Regionalidade*, v. 28, n. 82, p. 5–17, 2012. Citado 4 vezes nas páginas 19, 35, 36 e 70.

- FIGUEIREDO, M. A. D. *Sistema de medição de desempenho organizacional: um modelo para auxiliar a sua auto-avaliação*. Tese (Doutorado) — Universidade Federal do Rio de Janeiro - UFRJ, Rio de Janeiro, 2003. Citado 2 vezes nas páginas 54 e 55.
- FPNQ, F. p. o. P. N. d. Q. *Planejamento do Sistema de Medição de Desempenho Global*. São Paulo: [s.n.], 2001. Citado na página 54.
- FRANCISCHINI, A. S. N.; FRANCISCHINI, P. G. *Indicadores de Desempenho*. Rio de Janeiro: Alta Books, 2017. Citado 8 vezes nas páginas 20, 23, 52, 53, 70, 78, 79 e 88.
- FRETAS, C. D. S. et al. Introdução à visualização de informações. *RITA*, Rio de Janeiro, v. 8, n. 2, p. 143–158, 2001. Citado 2 vezes nas páginas 71 e 72.
- FUNDAÇÃO NACIONAL, d. Q. F. *Crêterios de excelência: Avaliação e diagnóstico da gestão organizacional*. São Paulo: [s.n.], 2007. Citado na página 54.
- FURTADO, J. C. et al. Ferramenta para extração de dados semiestruturados para carga de um big data. *Revista Brasileira de Computação Aplicada*, Passo Fundo, v. 7, n. 3, p. 43–52, 2015. Citado na página 61.
- GASQUES, J. G. et al. Desempenho e crescimento do agronegócio no brasil. *IPEA*, Brasília, 2004. Citado 8 vezes nas páginas 21, 56, 57, 70, 88, 95, 141 e 142.
- GHALAYINI, A. M.; NOBLE, J. S. The changing basis of performance measurement. *International Journal of Operations and Production Management*, v. 16, p. 63–80, 1996. Citado na página 54.
- GHALAYINI, A. M.; NOBLE, J. S.; CROWE, T. J. An integrated dynamic performance measurement system for improving manufacturing competitiveness. *International Journal Production Economics*, v. 48, p. 207–225, 1997. Citado na página 54.
- GIL, A. C. *Como Elaborar Projetos de Pesquisa*. São Paulo: [s.n.], 2002. v. 4. Citado na página 69.
- GIORDANO, S. R. Gestão ambiental no sistema agroindustrial. In: *Economia e gestão dos negócios agroalimentares*. [S.l.]: Pioneira, 2000. p. 235–254. Citado na página 36.
- HADDAWAY, N. R. The use of web-scraping software in searching for grey literature. *The Gray Journal*, v. 11, p. 186–190, 2015. Citado na página 65.
- HARRINGTON, H. J. *Aperfeiçoando processos empresariais*. São Paulo: Makron Books, 1993. Citado na página 52.
- HOUREAUX JUNIOR, F. *Avaliação de desempenho organizacional: estudo de casos de empresas do setor químico*. Dissertação (Mestrado) — Universidade de São Paulo - USP, São Paulo, 2005. Citado na página 54.
- HRONEC, S. M. *Sinais Vitais: usando medidas de desempenho da qualidade, tempo e custo para traçar a rota para o futuro de sua empresa*. São Paulo: [s.n.], 1994. Citado na página 54.
- IBGE, I. B. d. G. e. E. *Tendências Demográficas Uma análise dos resultados da Sinopse Preliminar do Censo Demográfico 2000*. Rio de Janeiro: IBGE, 2001. Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv773.pdf>>. Citado na página 36.

IBGE, I. B. d. G. e. E. *Censo Agropecuário*. Rio de Janeiro: IBGE, 2006. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/periodicos/51/agro_2006.pdf>. Citado na página 37.

IBGE, I. B. d. G. e. E. *Tendências Demográficas: uma análise da população com base nos resultados dos censos demográficos 1940 e 2000*. Rio de Janeiro: IBGE, 2007. Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv34956.pdf>>. Citado na página 35.

IBGE, I. B. d. G. e. E. *Censo Demográfico 2010*. Rio de Janeiro: IBGE, 2011. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/periodicos/93/cd_2010_caracteristicas_populacao_domicilios.pdf>. Citado na página 36.

INMON, W. H. *Building the Data Warehouse*. Indianápolis: [s.n.], 2005. v. 4. Citado na página 60.

JAESCHKE, G.; LEISLER, M.; HEMMJE, M. Modeling interactive, 3-dimensional information visualizations supporting information seeking behaviors. In: *Knowledge and Information Visualization: Searching for Synergies*. Berlim: Springer, 2005. p. 119–135. Citado na página 75.

JANK, M. S.; NASSAR, A. M. Competitividade e globalização. In: *Economia e gestão dos negócios agroalimentares*. [S.l.]: Pioneira, 2000. p. 138–164. Citado na página 49.

JSON.ORG. *Introdução ao JSON*. 2018. Disponível em: <<https://www.json.org/json-pt.html>>. Citado 2 vezes nas páginas 63 e 64.

JUNIOR, W. T. L.; OLIVEIRA, A. R. Jornalismo estruturado: uso de metadados para enriquecimento de bases noticiosas na web. *Revista Famecos*, Porto Alegre, n. 3, 2017. Citado na página 67.

KAPLAN, R. S.; NORTON, D. P. The balanced scorecard - measures that drive performance. *Harvard Business Review*, Boston, p. 71–79, 1992. Citado na página 54.

KARNOPP, E.; OLIVEIRA, V. S. Agronegócio e agricultura familiar: reflexões sobre sistemas produtivos do espaço agrário brasileiro. *Redes - Rev. Des. Regional*, Santa Cruz do Sul, v. 17, n. 2, p. 215–228, 2012. Citado 2 vezes nas páginas 19 e 70.

KHALIL, S.; FAKIR, M. Rcrawler: An r package for parallel web crawling and scraping. *SoftwareX*, Elsevier, v. 6, p. 98–106, 2017. Citado na página 66.

KIMBALL, R.; CASERTA, J. *The Data Warehouse ETL Toolkit Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. [S.l.: s.n.], 2004. Citado 3 vezes nas páginas 60, 61 e 63.

KOBRIN, S. J. et al. The assessment and evaluation of noneconomic environments by american firms: a preliminary report. *Palgrave Macmillan Journals*, n. 11, p. 32–47, 1980. Citado 2 vezes nas páginas 58 e 82.

KREUZ, C. L.; SOUZA, A.; CLEMENTE, A. Custos de produção, expectativas de retorno e de riscos do agronegócio mel no planalto norte de santa catarina. *XIV Congresso Brasileiro de Custos*, João Pessoa, 2007. Citado na página 56.

- LIMA, F. V. Territorialização do agronegócio e resistência camponesa. *Mercator*, Fortaleza, v. 15, n. 1, p. 73–86, 2016. Citado 3 vezes nas páginas 31, 32 e 35.
- LOHSE, G. L. et al. A classification of visual representations. *Visualization and Design*, v. 37, p. 36–49, 1994. Citado na página 75.
- LOUSADA, M.; VALENTIN, M. L. P. Modelos de tomada de decisão e sua relação com a informação orgânica. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 16, p. 147–164, 2011. Citado na página 43.
- LYNCH, R. L.; CROSS, K. F. *Measure Up!: Yardsticks for Continuous Improvement*. São Paulo: Blackwell Business, 1991. Citado na página 54.
- LÓPEZ, A. Pasos para elaborar el cuadro de mando. *Estrategia Financeira*, n. 149, p. 4–14, 1999. Citado na página 55.
- MACARTHUR, J. B. Performance measures that count: Monitoring variables of strategic importance. v. 10, p. 39–45, 1996. Citado na página 50.
- MACDONALD-ROSS, M. How numbers are shown: A review of research on the presentation of quantitative data in texts. *Springer*, v. 25, p. 359–409, 1977. Citado 2 vezes nas páginas 75 e 76.
- MAGALHÃES, J. et al. Extração e tratamento de dados na base lattes para identificação de core competencies em dengue. *Inf. Inf.*, Londrina, v. 19, n. 3, p. 30–54, 2014. Citado na página 61.
- MAIA, P. B. et al. Caracterização da produção de cana-de-açúcar através de indicadores espaciais e temporais em mato grosso do sul. *Revista Agrogeoambiental*, Porto Alegre, v. 7, p. 33–46, 2015. Citado 6 vezes nas páginas 21, 57, 70, 95, 131 e 132.
- MANOVICH, L. O que é visualização? *Estudos em Jornalismo e Mídia*, v. 8, p. 146–172, 2011. Citado na página 70.
- MARAVILHA-LOPES, S. P. A importância dos profissionais de gestão da informação para as organizações. *Biblios: Journal of Librarianship and Information Science*, Lima, v. 51, p. 91–98, 2013. Citado na página 42.
- MARCHAND, D. A.; KETTINGER, W. J.; ROLLINS, J. D. *Information Orientation: The Link to Business Performance*. Nova Iorque: [s.n.], 2002. v. 41. Citado 2 vezes nas páginas 47 e 48.
- MARCONI, M. d. A.; LAKATOS, E. M. *Fundamentos da metodologia científica*. São Paulo: [s.n.], 2007. v. 6. Citado 4 vezes nas páginas 27, 61, 69 e 85.
- MARTINDELL, J. *Harper and Brothers*. Nova Iorque: The scientific appraisal of management, 1950. Citado na página 54.
- MARTINS, M. A. Avaliação de desempenho empresarial como ferramenta para agregar valor ao negócio. *ConTexto*, Porto Alegre, v. 6, n. 10, p. 1–27, 2006. Citado na página 23.

MARTINS, R. A.; COSTA NETO, P. L. d. O. Indicadores de desempenho para a gestão pela qualidade total uma proposta de sistematização. *Gestão e Produção*, São Carlos, v. 5, p. 298–311, 1998. Citado 2 vezes nas páginas 50 e 52.

MAZOYER, M.; ROUDART, L. *História das agriculturas no mundo do neolítico à crise contemporânea*. São Paulo: Unesp, 2008. Citado 3 vezes nas páginas 31, 32 e 33.

MCKINNEY, W. *Python for Data Analysis*. 1. ed. [S.I.]: O'Reilly Media, Inc., 2012. Citado na página 65.

MEGIDO, J. L. T.; XAVIER, C. *Marketing e Agribusiness*. São Paulo: Atlas, 2003. v. 4. Citado 2 vezes nas páginas 20 e 37.

MELO, P. C. T.; VILELA, N. J. Desempenho da cadeia agroindustrial brasileira do tomate na década de 90. *Horticultura Brasileira*, Brasília, v. 22, p. 54–160, 2004. Citado 3 vezes nas páginas 56, 57 e 58.

MENA-CHALCO, J. P.; CESAR JUNIOR, R. M. Prospecção de dados acadêmicos de currículos lattes através do scriptlattes. *Pedro e João Editores*, São Carlos, p. 109–128, 2013. Citado na página 65.

MENDONÇA, M. L. O papel da agricultura nas relações internacionais e a construção do conceito de agronegócio. *Contexto Internacional*, Rio de Janeiro, v. 37, n. 2, p. 375–402, 2015. Citado 7 vezes nas páginas 19, 32, 33, 34, 35, 39 e 70.

MIRANDA, G. J.; REIS, E. A. Indicadores financeiros e não financeiros de longo prazo: um estudo em empresas atacadistas e distribuidoras brasileiras. *Contabilidade Vista e Revista*, Belo Horizonte, v. 17, n. 4, p. 11–34, 2006. Citado 2 vezes nas páginas 20 e 21.

MITCHELL, R. *Instant web scraping with Java*. Birmingham: Packt Publ., 2013. Citado na página 65.

MITCHELL, R. *Web Scraping with Python: Collecting Data from the Modern Web*. 1st. ed. [S.I.]: O'Reilly Media, Inc., 2015. Citado na página 65.

MONTALLI, K. M. L. Informação na indústria de bens de capital no brasil. *Ciência e Informação*, Brasília, n. 20, p. 45–50, 1991. Citado 2 vezes nas páginas 58 e 82.

MOON, K.-W. *Learn Ggplot2 Using Shiny App*. 1. ed. [S.I.]: Springer Publishing Company, Incorporated, 2017. Citado na página 66.

MOSLEY, M. *DAMA guide to the data management body of knowledge*. Technics Publications. Rio de Janeiro: Technics Publications, 2009. Citado na página 19.

MUNZERT, S. et al. *Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining*. 1st. ed. [S.I.]: Wiley Publishing, 2014. Citado 2 vezes nas páginas 65 e 66.

MÖLLER, H. D. Métodos quantitativos aplicados ao agronegócio. In: *Agronegócio*. São Paulo: Atlas, 2011. v. 3, p. 183–203. Citado 2 vezes nas páginas 49 e 50.

NAIR, V. G. *Getting Started with Beautiful Soup*. [S.I.]: Packt Publishing, 2014. Citado na página 65.

NAKAMURA, W. T.; MINETA, R. K. N. Identificação dos fatores que induzem ao uso do balanced scorecard como instrumento de gestão estratégica. In: *Encontro da Associação Nacional dos Programas de Pós-Graduação em Administração*. Campinas: ANPAD, 2001. Citado na página 53.

NASCIMENTO, H. A. D.; FERREIRA, C. B. R. Visualização de informações - uma abordagem prática. *XXV Congresso da Sociedade Brasileira de Computação*, São Leopoldo, p. 1262–1312, 2005. Citado 2 vezes nas páginas 71 e 78.

NASCIMENTO, S. et al. Mapeamento dos indicadores de desempenho organizacional em pesquisas da área de administração, ciências contábeis e turismo no período de 2000 a 2008. *Revista de Administração*, São Paulo, v. 46, n. 4, p. 373–391, 2011. Citado na página 50.

NEELY, A. The performance measurement revolution: why now and what next? *International Journal of Operations and Production Management*, v. 19, p. 205–228, 1999. Citado na página 50.

NEELY, A.; ADAMS, C.; CROWE, P. The performance prism in practice. *Measuring Business Excellence*, v. 5, p. 6 – 12, 2001. Citado na página 54.

NEELY, A.; GREGORY, M.; PLATTS, K. Performance measurement system design a literature review and research agenda. *International Journal of Operations and Production Management*, v. 15, p. 80–116, 1995. Citado 2 vezes nas páginas 53 e 54.

NELLI, F. *Python Data Analytics*. 1. ed. [S.l.]: Apress, 2015. Citado na página 65.

NILSSON, M. et al. Formalizing dublin core application profiles – description set profiles and graph constraints. In: *Metadata and Semantics*. Boston: Springer US, 2009. p. 101–111. Citado na página 67.

NUINTIN, A. A.; NAKAO, S. H. A definição de indicadores do desempenho e da qualidade para o processo de produção: estudo de casos do processo de produção do café. *Revista Contemporânea de Contabilidade*, Florianópolis, v. 7, p. 51–74, 2010. Citado na página 53.

NUTINI, M. A. *Transformando o Sistema de Indicadores: Avaliação do Desempenho Global sob a Ótica do MEG*. São Paulo: Fundação Nacional da Qualidade - FNQ, 2015. Citado 3 vezes nas páginas 23, 54 e 79.

OIKAWA, I. *Proposta de método para integração de bases de dados do agronegócio do Paraná*. Dissertação (Mestrado) — Universidade Federal do Paraná - UFPR, Curitiba, 2017. Citado 5 vezes nas páginas 20, 21, 23, 67 e 69.

OLIVEIRA, J. G. et al. Desempenho organizacional: a conexão entre quatro grandes disciplinas para ganho de competitividade. *CIKI - V Congresso Internacional de Conhecimento e Inovação*, Joinville, p. 1–23, 2015. Citado na página 47.

O'BRIEN, J. *Sistemas de informação e as decisões gerenciais na era da Internet*. São Paulo: Saraiva, 2004. Citado 4 vezes nas páginas 41, 44, 45 e 46.

PAIVA JUNIOR, F. G. Empreendedorismo e competencia do gestor no agronegócio. In: *Agronegócio*. São Paulo: Atlas, 2011. p. 48–57. Citado 2 vezes nas páginas 34 e 49.

PAULA, R. N. C. d.; ICHIKAWA, E. Y. Indicadores de produtividade em cooperativas do paran : um estudo comparativo de casos. In: *Encontro da Associa  o Nacional dos Programas de P s-Gradua  o em Administra  o*. Salvador: [s.n.], 2002. Citado 3 vezes nas p ginas 52, 55 e 56.

PEREIRA, F. C. M. Fontes de informa  o para neg cios: an lise sobre frequ ncia, relev ncia e confiabilidade, baseada em estudo emp rico com empres rios e gestores organizacionais. *Perspectivas em Ci ncia da Informa  o*, Belo Horizonte, v. 21, n. 2, p. 100–119, 2016. Citado na p gina 58.

PEREIRA, F. P. A. *Big Data e Data Analysis: Visualiza  o de Informa  o*. Disserta  o (Mestrado) — Universidade do Minho, Braga, 2015. Citado 3 vezes nas p ginas 76, 77 e 79.

PEREIRA, M. N. et al. Indicadores de desempenho de fazendas leiteiras de minas gerais. *Arquivo Brasileiro de Medicina Veterin ria e Zootecnia*, v. 68, n. 4, p. 1033–1042, 2016. Citado 4 vezes nas p ginas 56, 57, 58 e 70.

PETERS, R. *Expert Shell Scripting*. [S.l.]: Apress, 2009. Citado na p gina 90.

PETRI, S. M. *Modelo para apoiar a avalia  o das abordagens de gest o de desempenho e sugerir aperfei oamentos: sob a  tica construtivista*. Tese (Doutorado) — Universidade Federal de Santa Catarina, Florian polis, 2005. Citado na p gina 53.

PIKOR, P.; PIOTROWSKI, M. Data collection through web harvesting for real estate market research. In: *CREATIVITY SUPPORT SYSTEMS*. Katowice: Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, 2011. p. 113–122. Citado na p gina 105.

PINHEIRO, J. I. D. et al. *Probabilidade e Estat stica: Quantificando a incerteza*. Rio de Janeiro: Elsevier, 2012. Citado na p gina 50.

PINTO, R. C. Conflitos da era da informa  o: propriedade intelectual e cultura livre. *Amazon's Research and Environmental Law*, Ariquemes, v. 1, n. 1, p. 52–75, 2013. Citado na p gina 42.

PORTO, A.; ESTRADA, R. J. S. O uso de indicadores de desempenho como suporte   gest o estrat gica de uma ind stria moveleira – um estudo de caso. *XXIV Encontro Nac. de Eng. de Produ  o*, Florian polis, p. 3310–3317, 2004. Citado 2 vezes nas p ginas 50 e 51.

PYTHON.ORG. *General Python FAQ*. 2018. Dispon vel em: <<https://docs.python.org/3/faq/general.html>>. Citado na p gina 65.

QUESADO, P. R.; GUZM N, B. A.; RODRIGUES, L. L. O tableau de bord e o balanced scorecard : Uma an lise comparativa. *Revista de Contabilidade e Controladoria*, Curitiba, v. 4, n. 2, p. 128–150, 2012. Citado na p gina 55.

R CORE TEAM. *R Language Definition*. [S.l.], 2018. 60 p. Citado na p gina 65.

R, P. *What is R?* 2018. Dispon vel em: <<https://www.r-project.org/about.html>>. Citado na p gina 65.

RIBEMBOIM, J. A. Produtos agr colas e mercados no agroneg cio. In: *Agroneg cio*. S o Paulo: Atlas, 2011. p. 58–73. Citado na p gina 37.

RILEY, J. *Understanding Metadata: What is Metadata, and What is it For?* Baltimore: University of North Texas Libraries, 2017. Citado na página 67.

ROCHA JUNIOR, W. F.; BITTENCOURT, M. V. L.; RIBEIRO, M. C. P. Análise das características dos contratos no agronegócio do Brasil. *Revista Brasileira de Planejamento e Desenvolvimento*, Curitiba, v. 4, p. 94–118, 2015. Citado 3 vezes nas páginas 22, 36 e 70.

SANTOS, A. S.; MIRANDA, G. J. E-commerce no Brasil: se o segmento é promissor, por que empresas atuantes estão no vermelho? *Revista Evidenciação Contábil e Finanças*, João Pessoa, v. 3, p. 54–68, 2015. Citado na página 56.

SANTOS, C. D. Difusão do agronegócio e reestruturação urbano-regional no oeste baiano. *GeoTextos*, Salvador, v. 12, p. 157–181, 2016. Citado 2 vezes nas páginas 34 e 35.

SANTOS, G. J.; MARION, J. C.; SEGATTI, S. *Administração de Custos na Agropecuária*. São Paulo: Atlas, 2009. Citado na página 49.

SCHEREN, G.; WERNKE, R.; ZANIN, A. Proposta de indicadores financeiros sobre a judicialização da saúde no município de Chapecó (SC). *Revista Catarinense da Ciência Contábil*, Florianópolis, v. 17, n. 50, p. 106–123, 2018. Citado na página 57.

SETZER, W. W. Dado, informação, conhecimento e competência. *DataGramaZero - Revista de Ciência da Informação*, 1999. Citado na página 45.

SHAFRANOVICH, Y. *Common Format and MIME Type for Comma-Separated Values (CSV) Files*. The Internet Society, 2005. Disponível em: <<https://tools.ietf.org/pdf/rfc4180.pdf>>. Citado na página 63.

SHARDA, R.; DELEN, D.; EFRAIM, T. *Business Intelligence: A Managerial Perspective on Analytics*. Upper Saddle River: Pearson, 2013. v. 3. Citado na página 80.

SHNEIDERMAN, B. The eyes have it: A task by data type taxonomy for information visualizations. In: *Proceedings of the IEEE Symposium on Visual Languages*. [S.l.: s.n.], 1996. p. 336–343. Citado 3 vezes nas páginas 71, 73 e 74.

SICILIA, M.-A.; LYTRAS, M. D. *Metadata and Semantics*. 1. ed. [S.l.]: Springer Publishing Company, Incorporated, 2009. Citado na página 67.

SILVA, G. R. Gestão da informação para a tomada de decisão em uma instituição de ensino superior privada: A experiência das faculdades integradas da união educacional do planalto central (fapiplac/df). *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 15, p. 53–81, 2017. Citado na página 45.

SILVA NETO, M. A. *Mineração visual de dados: extração do conhecimento a partir das técnicas de visualização da informação e mineração de dados*. Dissertação (Mestrado) — Universidade Federal do Paraná - USP, Curitiba, 2008. Citado na página 78.

SILVA, R. J. Leitores e biblioteca escolar: do período neolítico ao *homo sapiens sapiens*. *Ensino Em Re-vista*, Uberlândia, v. 20, p. 357–366, 2013. Citado na página 31.

- SIMON, H. A. *The new science of management decision*. [S.l.: s.n.], 1960. v. 1. Citado na página 44.
- SINK, D. S.; TUTTLE, T. C. *Planejamento e medição para a performance*. Rio de Janeiro: Qualitymark, 1993. Citado 2 vezes nas páginas 52 e 54.
- SOARES, A. P. A. Teoria dos jogos, tomada de decisão e agronegócio. In: *Agronegócio*. São Paulo: Atlas, 2011. v. 3, p. 161–182. Citado 3 vezes nas páginas 43, 48 e 49.
- SOUSA, S. S. S.; BONJOUR, S. C. d. M.; FIGUEIREDO, A. M. R. análise da competitividade do algodão mato-grossense no cenário internacional: 1999 a 2005. *Revista de Estudos Sociais*, Cuiabá, p. 71–86, 2006. Citado 6 vezes nas páginas 21, 70, 95, 134, 135 e 136.
- SOUZA, M. A.; RASIA, K. A. Custos no agronegócio um perfil dos artigos publicados nos anais do congresso brasileiro de custos no período de 1998 a 2008. *Contabilidade, Gestao e Governanca*, Brasília, v. 14, n. 1, p. 69–81, 2011. Citado 2 vezes nas páginas 33 e 37.
- SOUZA, M. I. F.; VENDRUSCULO, L. G.; MELO, G. C. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão dublin core. *Ciência da Informação*, Brasília, v. 29, p. 93 – 102, 2000. Citado na página 67.
- SOUZA, U. B.; SANTOS, P. S.; BRANDÃO, P. R. B. Agronegócio e mudanças espaciais: um estudo de caso do espaço barreirense. *Caminhos de Geografia*, Uberlândia, v. 16, n. 56, p. 95–109, 2015. Citado na página 35.
- SULLIVAN, E. Optim: linking cost: time, and quality. *Quality Progress*, v. 19, n. 4, p. 52–53, 1986. Citado na página 54.
- SVEIBY, K. E. *A nova riqueza das organizações: gerenciando e avaliando patrimônios de conhecimento*. Rio de Janeiro: Campus, 2003. v. 7. Citado na página 42.
- TAFFAREL, M.; GARCIA PAULO MELLO, G. W. Sadia s.a. e perdigão s.a.: análise da trajetória histórica e do desempenho que antecederam o seu processo de fusão. *Revista Eletrônica de Ciência Administrativa*, Curitiba, v. 8, p. 59–69, 2009. Citado na página 57.
- TECH, A. R. B. *Desenvolvimento de uma ferramenta computacional para monitoramento e coleta de dados, baseado em conceitos de e-Science e Data Warehouse para aplicação na pecuária*. Tese (Doutorado) — Universidade de São Paulo - USP, Pirassununga, 2008. Citado na página 23.
- THEODOROU, P. Business strategy, structure and it alignment. In: *Encyclopedia of Information Science and Technology*. [S.l.: s.n.], 2005. p. 356–362. Citado na página 48.
- THOW-YIK, L. The basic entity model: a fundamental theorical model of information and information processing. *Informanon Processing and Management*, v. 30, p. 647–661, 1994. Citado na página 45.

- TOCCHETTO, M. R. L.; PEREIRA, L. C. Seleção de indicadores ambientais para indústrias com atividade galvânica. In: *Encontro da Associação Nacional de Pós-Graduação e Pesquisa em Administração*. Curitiba: [s.n.], 2008. Citado na página 50.
- TOLEDO FILHO, J. R. *Mercado de Capitais Brasileiro*. Campinas: Lucre, 1991. Citado na página 37.
- TUFTE, E. T. *Envisioning Information*. Cheshire: Graphics Press, 1990. Citado 2 vezes nas páginas 72 e 73.
- TUFTE, E. T. *The visual display of quantitative information*. Cheshire: Graphics Press, 2001. Citado na página 75.
- TUKEY, J. W. *Exploratory Data Analysis*. [S.l.]: Addison Wesley, 1977. Citado na página 50.
- TURLAND, M. *Php, Architect's Guide to Web Scraping*. [S.l.]: Marco Tabini & Associates, Inc., 2010. Citado 2 vezes nas páginas 65 e 67.
- VALIATI, E. R. d. A. *Avaliação de usabilidade de técnicas de visualização de informações multidimensionais*. Tese (Doutorado) — Universidade Federal do Rio grande do Sul, Porto Alegre, 2008. Citado na página 71.
- VAN BELLEN, H. M. Indicadores de sustentabilidade: um levantamento dos principais sistemas de avaliação. *Cadernos EBAPE.BR*, v. 2, p. 1 – 14, 2004. Citado na página 50.
- VARGIU, E.; URRU, M. Exploiting web scraping in a collaborative filtering based approach to web advertising. *Artif. Intell. Research*, v. 2, n. 1, p. 44–54, 2013. Citado na página 65.
- VERZANI, J. *Getting Started with RStudio*. [S.l.]: O'Reilly, 2011. Citado 2 vezes nas páginas 65 e 82.
- W3C, W. W. W. C. *Extensible Markup Language (XML)*. [S.l.]: World Wide Web Consortium, 2006. Citado na página 63.
- WAACK, R. S. Gerenciamento de tecnologia e inovação em sistemas agroindustriais. In: *Economia e gestão dos negócios agroalimentares*. [S.l.]: Pioneira, 2000. p. 323–348. Citado na página 48.
- WARD, J. *Instant PHP Web Scraping*. [S.l.]: Packt Publishing, 2013. Citado na página 67.
- WARD, J.; PEPPARD, J. *Strategic Planning for Information Systems*. Chichester: John Wiley Sons, Inc, 2002. Citado 2 vezes nas páginas 46 e 48.
- WARE, C. *Information Visualization: Perception for Design*. Campinas: Morgan Kaufmann, 2004. Citado na página 73.
- WEIBEL, S. et al. *Dublin Core Metadata for Resource Discovery*. United States: RFC Editor, 1998. Citado 2 vezes nas páginas 67 e 68.

WELCH, C. A.; FERNANDES, B. M. Agricultura e mercado: campesinato e agronegócio da laranja nos eua e brasil. In: *Campesinato e territórios em disputa*. São Paulo: Expressão Popular, 2008. p. 161–190. Citado na página 20.

WICKHAM, H. *ggplot2 - Elegant graphics for data analysis*. New York: [s.n.], 2009. Citado na página 66.

WICKHAM, H. Tidy data. *Journal of Statistical Software*, v. 59, n. 10, p. 1–17, 2014. Citado 2 vezes nas páginas 61 e 62.

WICKHAM, H. et al. dplyr: A grammar of data manipulation. *R package version 0.4*, v. 3, 2015. Citado na página 66.

WICKHAM, H.; GROLEMUND, G. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 1. ed. [S.l.]: O'Reilly Media, Inc., 2017. Citado 3 vezes nas páginas 62, 66 e 67.

WILDAUER, E. W. Análise de dados e formulação de indicadores de produção do agronegócio do estado do paran . 2015. Citado na p gina 20.

YI, J. S. et al. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, Piscataway, v. 13, n. 6, p. 1224–1231, 2007. Citado 2 vezes nas p ginas 73 e 74.

YOUNG, C. S. Tecnologia de informa  o. *Revista de Administra  o de Empresas*, S o Paulo, v. 32, p. 78–87, 1992. Citado na p gina 48.

ZYLBERSZTAJN, D. *Economia e Gest o dos Neg cios Agroalimentares*. S o Paulo: Pioneira, 2000. Citado 2 vezes nas p ginas 20 e 37.

APÊNDICE A – DADOS DE PRODUTOS DISPONIBILIZADOS PELA CONAB

QUADRO 13 – DADOS DA AGRICULTURA DISPONIBILIZADOS PELA CONAB

Dados de produção, área plantada, produtividade - 1976 - 2018					
1	Cana-de-Açúcar - Agrícola	2	Algodão	3	Amendoim 1ª Safra
4	Amendoim 2ª Safra	5	Amendoim Total (1ª e 2ª Safras)	6	Arroz Irrigado
7	Arroz Sequeiro	8	Arroz Total (Irrigado e Sequeiro)	9	Aveia
10	Canola	11	Centeio	12	Cevada
13	Feijão 1ª Safra	14	Feijão 2ª Safra	15	Feijão 3ª Safra
16	Feijão Total (1ª, 2ª e 3ª Safras)	17	Girassol	18	Mamona
19	Milho 1ª Safra	20	Milho 2ª Safra	21	Milho Total (1ª e 2ª Safras)
22	Soja	23	Sorgo	24	Trigo
25	Triticale				

inputapendices/dados_{embrapa}.tex

APÊNDICE B – DADOS DE PRODUTOS DISPONIBILIZADOS PELO IBGE

QUADRO 14 – DADOS DE AGRICULTURA DISPONIBILIZADOS PELO IBGE

Dados de produção, produtividade, área planta e colhida - 2006 - 2018					
1	Abacaxi	2	Algodão herbáceo	3	Alho
4	Amendoim (1ª Safra)	5	Amendoim (2ª Safra)	6	Arroz
7	Aveia	8	Banana	9	Batata - inglesa (1ª Safra)
10	Batata - inglesa (2ª Safra)	11	Batata - inglesa (3ª Safra)	12	Cacau
13	Café arábica	14	Café canephora	15	Café total
16	Cana-de-açúcar	17	Castanha-de-caju	18	Cebola
19	Centeio	20	Cereais, leguminosas e oleaginosas	21	Cevada
22	Coco-da-baía	23	Feijão (1ª Safra)	24	Feijão (2ª Safra)
25	Feijão (3ª Safra)	26	Fumo	27	Girassol
28	Guaraná	29	Juta	30	Laranja
31	Maçã	32	Malva	33	Mamona
34	Mandioca	35	Milho (1ª Safra)	36	Milho (2ª Safra)
37	Pimenta-do-reino	38	Sisal ou agave	39	Soja
40	Sorgo	41	Tomate	42	Trigo
43	Triticale	44	Uva		

QUADRO 15 – DADOS DE PECUÁRIA DISPONIBILIZADOS PELO IBGE - ABATE

Dados de abate e peso total das carcaças - 1997 - 2018					
1	Bois	2	Frangos	3	Novilhas
4	Novilhos	5	Suínos	6	Vacas
7	Vitelos e Vitelas				

APÊNDICE C – DADOS DE PRODUTOS DISPONIBILIZADOS PELA FAO

QUADRO 16 – DADOS DA AGRICULTURA DISPONIBILIZADOS PELA FAO

Dados de produção, produtividade e área colhida - 1961 - 2016					
1	Abacate	2	Abacaxi	3	Algodão
4	Alho	5	Amendoim com Casca	6	Arroz
7	Aveia	8	Azeitona	9	Banana
10	Batata	11	Batata-doce	12	Borracha
13	Café Verde	14	Caju	15	Cana-de-Açúcar
16	Caqui	17	Castanha	18	Castanha de Caju
19	Castanha-do-Pará	20	Cebola	21	Centeio
22	Cevada	23	Chá	24	Côco
25	Colza	26	Ervilha Seca	27	Feijão Cavalo
28	Feijão Seco	29	Fibra de Algodão	30	Fibra de Bast
31	Fibras	32	Figo	33	Fruta de Óleo de Palma
34	Frutas Tropicais	35	Goma	36	Grão de Palma
37	Inhame	38	Juta	39	Laranja
40	Legumes Frescos	41	Limão	42	Linhaça
43	Maçã	44	Mamão	45	Mandioca
46	Manga, Mangostão, Goiaba	47	Marmelo	48	Mate
49	Melancia	50	Melão	51	Milho
52	Morango	53	Noz	54	Noz com Casca
55	Noz de Tungue	56	Oleaginosas	57	Óleo de Palma
58	Pêra	59	Pêssego e Nectarina	60	Pimenta
61	Píreto	62	Rami	63	Semente de Algodão
64	Semente de Gergilim	65	Semente de Girassol	66	Semente de mamona
67	Sementes de cacau	68	Sisal	69	Soja
70	Sorgo	71	Tabaco	72	Tangerina, Clementina e Satsumo
73	Tomate	74	Toranja	75	Trigo
76	Trigo Mourisco	77	Triticale	78	Uva

QUADRO 17 – DADOS DE PECUÁRIA DISPONIBILIZADOS PELA FAO - QUANTIDADE DE ANIMAIS VIVOS

Dados de tamanho de rebanhos - 1961 - 2016					
1	Asno	2	Búfalo	3	Cabra
4	Cavalo	5	Coelho e Lebra	6	Colméia
7	Frango	8	Gado	9	Mula
10	Ovelha	11	Pato	12	Peru
13	Porco				

QUADRO 18 – DADOS DE PECUÁRIA DISPONIBILIZADOS PELA FAO - PRODUTOS PRIMÁRIOS DE ORIGEM ANIMAL

Dados de quantidade de animais abatidos, produção e produtividade - 1961 - 2016					
1	Bicho da Seda	2	Carne de Cabra	3	Carne de Cabra Selvagem
4	Carne de Cavalo	5	Carne de Cavalo Selvagem	6	Carne de Coelho
7	Carne de Coelho Selvagem	8	Carne de Frango	9	Carne de Frango Selvagem
10	Carne de Gado	11	Carne de Gado Selvagem	12	Carne de Ovelha
13	Carne de Ovelha Selvagem	14	Carne de Pato	15	Carne de Pato Selvagem
16	Carne de Peru	17	Carne de Peru Selvagem	18	Carne de Porco
19	Carne de Porco Selvagem	20	Cera de Abelha	21	Couro de Cabra
22	Couro de Gado	23	Couro de Ovelhas	24	Lã
25	Ovos de Galinha com Casca	26	Ovos de Outras Aves com Casca		

APÊNDICE D – SCRIPT DE COLETA E TRANSFORMAÇÃO DOS DADOS

```

1  install.packages("devtools", repos="http://cran.rstudio.com/")
2  library(devtools)
3  install.packages("xml2", repos="http://cran.rstudio.com/")
4  install.packages("rvest", repos="http://cran.rstudio.com/")
5  install.packages("tidyverse", repos="http://cran.rstudio.com/")
6  install.packages("readxl", repos="http://cran.rstudio.com/")
7  install.packages("rjson", repos="http://cran.rstudio.com/")
8  devtools::install_version("readxl", version = "1.0.0",
    repos="http://cran.rstudio.com/")
9  install.packages("writexl", repos="http://cran.rstudio.com/")
10 install.packages('rsconnect', repos="http://cran.rstudio.com/")
11
12 library(xml2)
13 library(rvest)
14 library(readxl)
15 library(tidyverse)
16 library(rjson)
17 library(writexl)
18 library(rsconnect)
19
20 #criando data frames e caminho para salvar arquivos
21 dadosAgricultura <- data.frame(1,2,3,4,5,6,7)
22 colnames(dadosAgricultura) <- c("ANO", "ESTADO", "FONTE",
    "PRODUTO", "INDICADOR", "UNIDADE", "VALOR")
23 dadosPecuararia <- data.frame(1,2,3,4,5,6,7)
24 colnames(dadosPecuararia) <- c("ANO", "ESTADO", "FONTE", "PRODUTO", "INDICADOR",
    "UNIDADE", "VALOR")
25 caminhoDownloadArquivos = "/home/jhonny/Downloads/"
26 #criando pastas onde ficarao salvo os arquivos de origem
27 dir.create(paste0(caminhoDownloadArquivos, "IBGE"),
28           showWarnings = TRUE,
29           recursive = FALSE, mode = "0777")
30 dir.create(paste0(caminhoDownloadArquivos, "CONAB"),
31           showWarnings = TRUE,
32           recursive = FALSE, mode = "0777")
33 dir.create(paste0(caminhoDownloadArquivos, "EMBRAPA"),

```

```
34         showWarnings = TRUE,
35         recursive = FALSE, mode = "0777")
36 dir.create(paste0(caminhoDownloadArquivos, "EMBRAPA/Agricultura"),
37           showWarnings = TRUE,
38           recursive = FALSE, mode = "0777")
39 dir.create(paste0(caminhoDownloadArquivos, "EMBRAPA/Pecuaria"),
40           showWarnings = TRUE,
41           recursive = FALSE, mode = "0777")
42 #conectando com shinyapps
43 rsconnect::setAccountInfo(name='nomeConta',
44   token='token',
45   secret='secret')
46
47 ##### dados da IBGE #####
48
49 #funcao de limpeza dos dados de agricultura
50 tryCatch({
51   limpaDadosAgricultura <- function(dadosAgricultura, nomeArquivo) {
52     abas = c("rea plantada (Hectares)", "rea colhida (Hectares)", "Produto
53             (Toneladas)", "Rendimento mdio (Quilograma...)");
54     for (numArquivo in 1:length(nomeArquivo)){
55       for(aba in 1:length(abas)){
56         dados <- data.frame(1,2,3,4,5,6,7)
57         colnames(dados) <- c("ANO", "ESTADO", "FONTE", "PRODUTO", "INDICADOR",
58                             "UNIDADE", "VALOR")
59         #abrindo arquivo salvo
60         produto <- read_excel(paste0("~/Downloads/IBGE/",
61                                     nomeArquivo[numArquivo]), sheet = abas[aba])
62
63         colnames(produto)[1] = "local"
64         colnames(produto)[2] = "nomeproduto"
65         #definindo o nome do produto
66         arrayNomeProduto = unlist(strsplit(produto$nomeproduto[4], " "))
67         nomeProduto = arrayNomeProduto[2]
68         if(length(arrayNomeProduto) > 2){
69           for(i in 3:length(arrayNomeProduto)){
70             nomeProduto = paste(nomeProduto, arrayNomeProduto[i], sep = " ")
71           }
72         }
73       }
74     }
75     #definindo indicador e unidade
```

```

72     indicadorUnidade =
          unlist(strsplit(stringr::str_replace_all(produto$local[1], "Varivel -
            ", " "), " \\("))
73     indicador = indicadorUnidade[[1]]
74     unidade = stringr::str_replace_all(indicadorUnidade[2], "\\)", "")
75
76     #renomeando as colunas para para depois vira-las
77     produto <- produto[-c(1, 2), ]
78     totalColunas = ncol(produto)
79     for(j in 2:totalColunas){
80         colnames(produto)[j] = paste0("periodo - ", produto[1,j])
81     }
82
83     produto <- produto %>% gather(periodo, valor,
          dplyr::starts_with("periodo"))
84     produto <- na.omit(produto)
85     produto["ANO"] <- c(0)
86     produto$ANO <- as.integer(substr(produto$periodo,
          nchar(produto$periodo)-4, nchar(produto$periodo)))
87     anoMaximo = as.integer(max(unique(produto$ANO)))
88     produto <- produto %>%
89         separate(periodo, c("periodo", "mes"), sep = " - ") %>%
90         separate(mes, c("mes", "ano"), sep = " ") %>%
91         mutate(numeroMes = case_when(mes == "janeiro" ~ 1,
92             mes == "fevereiro" ~ 2, mes == "maro" ~ 3, mes == "abril" ~ 4,
93             mes == "maio" ~ 5, mes == "junho" ~ 6, mes == "julho" ~ 7,
94             mes == "agosto" ~ 8, mes == "setembro" ~ 9,
95             mes == "outubro" ~ 10, mes == "novembro" ~ 11,
96             mes == "dezembro" ~ 12)) %>%
97         group_by(ano) %>% filter(numeroMes == max(numeroMes))
98
99     totalLinhas = nrow(produto)
100
101     #adicionando registros no data frame
102     for(i in 1:totalLinhas){
103         dados[nrow(dados)+1, ] <- data.frame(ANO = produto$ano[i],
104             ESTADO = produto$local[i],
105             FONTE = "IBGE",
106             PRODUTO = nomeProduto,
107             INDICADOR = indicador,
108             UNIDADE = unidade,
109             VALOR = produto$valor[i],

```

```
110         stringsAsFactors =FALSE)
111     }
112
113     dados <- mutate_each(dados, funs(as.numeric), VALOR)
114     dados <- dados %>% filter(ANO > 1)
115     anos = unique(dados$ANO);
116
117     #gerando registros com somatorios por regioa a cada ano
118     for(i in 1:length(anos)){
119         #fazendo somatorio dos dados da regioa sul
120         totalSul = sum(dados %>%
121             filter(ANO == anos[i] &
122                 PRODUTO == nomeProduto &
123                 INDICADOR == indicador &
124                 (ESTADO == "Paran" | ESTADO == "Rio Grande do Sul" |
125                 ESTADO == "Santa Catarina")) %>%
126             select(VALOR));
127         dados[nrow(dados)+1, ] <- data.frame(ANO = anos[i],ESTADO = "Sul",
128             FONTE = "IBGE",
129             PRODUTO = nomeProduto,
130             INDICADOR =indicador,
131             UNIDADE = unidade,
132             VALOR = totalSul,
133             stringsAsFactors =FALSE)
134         #fazendo somatorio dos dados da regioa norte
135         totalNorte = sum(dados %>%
136             filter(ANO == anos[i] &
137                 PRODUTO == nomeProduto
138                 & INDICADOR == indicador &
139                 (ESTADO == "Rondnia" | ESTADO == "Acre" |
140                 ESTADO == "Amazonas" | ESTADO == "Par" |
141                 ESTADO == "Roraima" | ESTADO == "Amap" |
142                 ESTADO == "Tocantins")) %>%
143             select(VALOR));
144         dados[nrow(dados)+1, ] <- data.frame(ANO = anos[i],ESTADO = "Norte",
145             FONTE = "IBGE",
146             PRODUTO = nomeProduto,
147             INDICADOR =indicador,
148             UNIDADE = unidade,
149             VALOR = totalNorte,
150             stringsAsFactors =FALSE)
151         #fazendo somatorio dos dados da regioa nordeste
```

```
152     totalNordeste = sum(dados %>%
153         filter(ANO == anos[i] &
154             PRODUTO == nomeProduto &
155             INDICADOR == indicador &
156             (ESTADO == "Maranhão" | ESTADO == "Piau" |
157             ESTADO == "Cear" | ESTADO == "Bahia" |
158             ESTADO == "Rio Grande do Norte" |
159             ESTADO == "Paraba" | ESTADO == "Pernambuco" |
160             ESTADO == "Alagoas" | ESTADO == "Sergipe"))
161         %>%
162         select(VALOR));
163     dados[nrow(dados)+1, ] <- data.frame(ANO = anos[i], ESTADO = "Nordeste",
164         FONTE = "IBGE",
165         PRODUTO = nomeProduto,
166         INDICADOR = indicador,
167         UNIDADE = unidade,
168         VALOR = totalNordeste,
169         stringsAsFactors = FALSE)
170     #fazendo somatorio dos dados da regioao centro-oeste
171     totalCentroOeste = sum(dados %>%
172         filter(ANO == anos[i] &
173             PRODUTO == nomeProduto &
174             INDICADOR == indicador &
175             (ESTADO == "Mato Grosso" |
176             ESTADO == "Mato Grosso do Sul" |
177             ESTADO == "Gois" |
178             ESTADO == "Distrito Federal")) %>%
179         select(VALOR));
180     dados[nrow(dados)+1, ] <- data.frame(ANO = anos[i], ESTADO =
181         "Centro-Oeste",
182         FONTE = "IBGE",
183         PRODUTO = nomeProduto,
184         INDICADOR = indicador,
185         UNIDADE = unidade,
186         VALOR = totalCentroOeste,
187         stringsAsFactors = FALSE)
188     #fazendo somatorio dos dados da regioao sudeste
189     totalSudeste = sum(dados %>%
190         filter(ANO == anos[i] &
191             PRODUTO == nomeProduto &
192             INDICADOR == indicador &
193             (ESTADO == "Rio de Janeiro" |
```

```
192         ESTADO == "So Paulo" |
193         ESTADO == "Esprito Santo" |
194         ESTADO == "Minas Gerais")) %>%
195         select(VALOR));
196     dados[nrow(dados)+1, ] <- data.frame(ANO = anos[i], ESTADO = "Sudeste",
197         FONTE = "IBGE",
198         PRODUTO = nomeProduto,
199         INDICADOR = indicador,
200         UNIDADE = unidade,
201         VALOR = totalSudeste,
202         stringsAsFactors = FALSE)
203 }
204 dadosAgricultura <- rbind(dadosAgricultura, dados)
205 }
206 }
207 return(dadosAgricultura);
208 }
209
210 #funcao de limpeza dos dados de pecuria
211 limpaDadosPecuarria <- function(dadosPecuarria, nomeArquivosPecuarria) {
212     numeroRegistroPecuarria = 1;
213     for (numArquivo in 1:length(nomeArquivosPecuarria)){
214         abas <- readxl::excel_sheets(paste0("~/Downloads/IBGE/",
215             nomeArquivosPecuarria[numArquivo]))
216         #a ultima aba nao deve ser vista, pois conta apenas com notas
217         for(numAba in 1:(length(abas)-1)){
218             dados <- data.frame(1,2,3,4,5,6,7)
219             colnames(dados) <- c("ANO", "ESTADO", "FONTE", "PRODUTO", "INDICADOR",
220                 "UNIDADE", "VALOR")
221             produto <- read_excel(paste0(caminhoDownloadArquivos, "IBGE/",
222                 nomeArquivosPecuarria[numArquivo]), sheet=abas[numAba])
223
224             #removendo a coluna 2
225             produto <- produto[, -(2)]
226             colnames(produto)[1] = "local"
227             colnames(produto)[2] = "nomeproduto"
228
229             #definindo o nome do produto
230             nomeProduto = produto$nomeproduto[5]
231             if(!grepl('bovinos', nomeArquivosPecuarria[numArquivo])){
232                 nomeProduto = nomeArquivosPecuarria[numArquivo]
233                 switch(nomeArquivosPecuarria[numArquivo],
```



```
231         abate_frangos.xlsx={nomeProduto = 'Frangos'},
232         abate_suinos.xlsx={nomeProduto = 'Sunos'},
233         producao_ovos.xlsx={nomeProduto = 'Ovos de Galinha'},
234         producao_leite.xlsx={nomeProduto = 'Leite de Vaca'}
235     )
236 }
237
238 #definindo indicador e unidade
239 indicadorUnidade =
240     unlist(strsplit(stringr::str_replace_all(produto$local[1], "Varivel -", " "), "\\\("))
241
242 indicador = indicadorUnidade[[1]]
243 unidade = stringr::str_replace_all(indicadorUnidade[2], "\\)", "")
244
245 #renomeando as colunas para para depois vira-las
246 produto <- produto[-c(1, 2), ]
247 totalColunas = ncol(produto)
248 for(j in 2:totalColunas){
249     colnames(produto)[j] = paste0("periodo - ", produto[1,j])
250 }
251 produto <- produto %>% gather(periodo, valor,
252     dplyr::starts_with("periodo"))
253 produto["ANO"] <- c(0)
254 produto["VALOR"] <- as.numeric(produto$valor)
255 produto <- produto[,-(3)]
256 produto["TOTALANO"] <- c(0)
257 produto$ANO <- as.integer(substr(produto$periodo,
258     nchar(produto$periodo)-4, nchar(produto$periodo)))
259
260 produto$VALOR[is.na(produto$VALOR)] <- 0
261 produto <- na.omit(produto)
262 anos = unique(produto$ANO);
263 estados = unique(produto$local)
264
265 #adicionando registros no data frame
266 for(i in 1:length(anos)){
267     for(j in 1:length(estados)){
268         total = sum(produto %>%
269             filter(ANO == anos[i] & local == estados[j]) %>%
270             select(VALOR));
271         dados[nrow(dados)+1, ] <- data.frame(ANO = anos[i],
272             ESTADO = estados[j],
```

```
269             FONTE = "IBGE",
270             PRODUTO = nomeProduto,
271             INDICADOR = indicador,
272             UNIDADE = unidade,
273             VALOR = total,
274             stringsAsFactors = FALSE)
275     }
276 }
277
278 #gerando registros com somatorios por regioa a cada ano
279 for(i in 1:length(anos)){
280     #fazendo somatorio dos dados da regioa sul
281     totalSul = sum(dados %>%
282         filter(ANO == anos[i] & PRODUTO == nomeProduto &
283             INDICADOR == indicador &
284             (ESTADO == "Paran" | ESTADO == "Rio Grande do Sul" |
285             ESTADO == "Santa Catarina")) %>%
286         select(VALOR));
287     dados[nrow(dados)+1, ] <- data.frame(ANO = anos[i],
288         ESTADO = "Sul",
289         FONTE = "IBGE",
290         PRODUTO = nomeProduto,
291         INDICADOR = indicador,
292         UNIDADE = unidade,
293         VALOR = totalSul,
294         stringsAsFactors = FALSE)
295     #fazendo somatorio dos dados da regioa norte
296     totalNorte = sum(dados %>%
297         filter(ANO == anos[i] &
298             PRODUTO == nomeProduto &
299             INDICADOR == indicador &
300             (ESTADO == "Rondnia" | ESTADO == "Acre" |
301             ESTADO == "Amazonas" | ESTADO == "Par" |
302             ESTADO == "Roraima" | ESTADO == "Amap" |
303             ESTADO == "Tocantins")) %>%
304         select(VALOR));
305     dados[nrow(dados)+1, ] <- data.frame(ANO = anos[i],
306         ESTADO = "Norte",
307         FONTE = "IBGE",
308         PRODUTO = nomeProduto,
309         INDICADOR = indicador,
310         UNIDADE = unidade,
```

```
311             VALOR = totalNorte,
312             stringsAsFactors =FALSE)
313     #fazendo somatorio dos dados da regioao nordeste
314     totalNordeste = sum(dados %>%
315         filter(ANO == anos[i] &
316             PRODUTO == nomeProduto &
317             INDICADOR == indicador &
318             (ESTADO == "Maranhao" | ESTADO == "Piau" |
319             ESTADO == "Cear" | ESTADO == "Bahia" |
320             ESTADO == "Rio Grande do Norte" |
321             ESTADO == "Paraba" | ESTADO == "Pernambuco" |
322             ESTADO == "Alagoas" |
323             ESTADO == "Sergipe"))) %>%
324         select(VALOR));
325     dados[nrow(dados)+1, ] <- data.frame(ANO = anos[i],
326         ESTADO = "Nordeste",
327         FONTE = "IBGE",
328         PRODUTO = nomeProduto,
329         INDICADOR = indicador,
330         UNIDADE = unidade,
331         VALOR = totalNordeste,
332         stringsAsFactors =FALSE)
333     #fazendo somatorio dos dados da regioao centro-oeste
334     totalCentroOeste = sum(dados %>%
335         filter(ANO == anos[i] &
336             PRODUTO == nomeProduto &
337             INDICADOR == indicador &
338             (ESTADO == "Mato Grosso" |
339             ESTADO == "Mato Grosso do Sul" |
340             ESTADO == "Gois" |
341             ESTADO == "Distrito Federal"))) %>%
342         select(VALOR));
343     dados[nrow(dados)+1, ] <- data.frame(ANO = anos[i],
344         ESTADO = "Centro-Oeste",
345         FONTE = "IBGE",
346         PRODUTO = nomeProduto,
347         INDICADOR = indicador,
348         UNIDADE = unidade,
349         VALOR = totalCentroOeste,
350         stringsAsFactors =FALSE)
351     #fazendo somatorio dos dados da regioao sudeste
352     totalSudeste = sum(dados %>%
```

```

353         filter(ANO == anos[i] &
354               PRODUTO == nomeProduto &
355               INDICADOR == indicador &
356               (ESTADO == "Rio de Janeiro" |
357               ESTADO == "So Paulo" |
358               ESTADO == "Esprito Santo" |
359               ESTADO == "Minas Gerais")) %>%
360         select(VLOR));
361     dados[nrow(dados)+1, ] <- data.frame(ANO = anos[i],
362                                         ESTADO = "Sudeste",
363                                         FONTE = "IBGE",
364                                         PRODUTO = nomeProduto,
365                                         INDICADOR = indicador,
366                                         UNIDADE = unidade,
367                                         VALOR = totalSudeste,
368                                         stringsAsFactors = FALSE)
369 }
370 dadosPecuaria <- rbind(dadosPecuaria, dados)
371 }
372 }
373 return(dadosPecuaria)
374 }
375 #funcao que baixa os arquivos a partir dos links
376 baixarPlanilhaIBGE <- function(nomeArquivos, linkArquivos) {
377   for (j in 1:length(nomeArquivos)){
378     caminhoDestino <- paste(caminhoDownloadArquivos, "IBGE/", nomeArquivos[j],
379                             sep="");
380     url <- paste("https://sidra.ibge.gov.br/geratabela?format=xlsx&",
381                 linkArquivos[j], sep="");
382     download.file(url, caminhoDestino);
383   }
384 }
385 #coletando dados da aba 1: abate de bovinos, sunos e frangos
386 nomesArquivosAbate <- c('abate_bovinos_bois.xlsx', 'abate_bovinos_vacas.xlsx',
387                         'abate_bovinos_novilhos.xlsx', 'abate_bovinos_novilhas.xlsx',
388                         'abate_bovinos_vitelos.xlsx', 'abate_suinos.xlsx',
389                         'abate_frangos.xlsx')
390 linksArquivosAbate <-
391   c('name=tabela1092.xlsx&terr=N&rank=-&query=t/1092/n1/all/
392     n3/all/v/284,285/p/all/c12716/115236/c18/55/c12529/118225/1/
393     v,p%2Bc12716%2Bc18,t%2Bc12529',
394     'name=tabela1092.xlsx&terr=N&rank=-&query=t/1092/n1/all/n3/all/

```

```

392     v/284,285/p/all/c12716/115236/c18/56/c12529/118225/l/
393     v,p%2Bc12716%2Bc18,t%2Bc12529',
394     'name=tabela1092.xlsx&terr=N&rank=-&query=t/1092/n1/all/n3/all/
395     v/284,285/p/all/c12716/115236/c18/111734/c12529/118225/l/
396     v,p%2Bc12716%2Bc18,t%2Bc12529',
397     'name=tabela1092.xlsx&terr=N&rank=-&query=t/1092/n1/all/n3/all/
398     v/284,285/p/all/c12716/115236/c18/111735/c12529/118225/l/
399     v,p%2Bc12716%2Bc18,t%2Bc12529',
400     'name=tabela1092.xlsx&terr=N&rank=-&query=t/1092/n1/all/n3/
401     all/v/284,285/p/all/c12716/115236/c18/57/c12529/118225/l/
402     v,p%2Bc12716%2Bc18,t%2Bc12529',
403     'name=tabela1093.xlsx&terr=N&rank=-&query=t/1093/n1/all/n3/
404     all/v/284,285/p/all/c12716/115236/c12529/118225/l/
405     v,p%2Bc12716,t%2Bc12529',
406     'name=tabela1094.xlsx&terr=N&rank=-&query=t/1094/n1/all/n3/
407     all/v/284,285/p/all/c12716/115236/c12529/118225/l/
408     v,p%2Bc12716,t%2Bc12529')
409 #baixando planilhas
410 baixarPlanilhaIBGE(nomesArquivosAbate, linksArquivosAbate);
411
412 #coletando dados da aba 2: producao de leite
413 nomeArquivoLeite <- 'producao_leite.xlsx'
414 linkArquivosLeite <-
415     'name=tabela1086.xlsx&terr=N&rank=-&query=t/1086/n1/
416     all/n3/all/v/282/p/all/c12716/115236/c12529/118225/d/v282%200/l/
417     v,p%2Bc12716,t%2Bc12529'
418
419 #baixando planilha
420 baixarPlanilhaIBGE(nomeArquivoLeite, linkArquivosLeite);
421
422 #coletando dados da aba 3: dados de producao de ovos
423 nomeArquivoOvo <- 'producao_ovos.xlsx'
424 linkArquivosOvo <-
425     'name=tabela915.xlsx&terr=N&rank=-&query=t/915/n1/all/n3/
426     all/v/29,1988/p/all/c12716/115236/l/
427     v,p%2Bc12716,t'
428
429 #baixando planilha
430 baixarPlanilhaIBGE(nomeArquivoOvo, linkArquivosOvo);
431
432 nomeArquivosPecuararia <- c(nomesArquivosAbate, nomeArquivoLeite, nomeArquivoOvo)
433

```

```
434 #limpando dados dos arquivos baixados
435 dadosPecuararia = limpaDadosPecuararia(dadosPecuararia, nomeArquivosPecuararia)
436
437 #coletando dados da aba 4: producao de safra
438 nomesArquivosAgricola <-
439   c('cereal_leguminosa_oleaginosa.xlsx', 'algodao_herbaceo.xlsx',
440     'amendoim1.xlsx', 'amendoim2.xlsx', 'arroz.xlsx', 'aveia.xlsx', 'centeio.xlsx',
441     'cevada.xlsx', 'feijao1.xlsx', 'feijao2.xlsx', 'feijao3.xlsx', 'girassol.xlsx',
442     'mamona.xlsx', 'milho1.xlsx', 'milho2.xlsx', 'soja.xlsx', 'sorgo.xlsx', 'trigo.xlsx',
443     'triticale.xlsx', 'abacaxi.xlsx', 'alho.xlsx', 'banana.xlsx', 'batata_inglesa1.xlsx',
444     'batata_inglesa2.xlsx', 'batata_inglesa3.xlsx', 'cacau.xlsx', 'cafe_arabica.xlsx',
445     'cafe_canephora.xlsx', 'cana_de_acucar.xlsx', 'castanha_de_caju.xlsx', 'cebola.xlsx',
446     'coco_da_baia.xlsx', 'fumo.xlsx', 'guarana.xlsx', 'juta.xlsx', 'laranja.xlsx',
447     'maca.xlsx', 'malva.xlsx', 'mandioca.xlsx', 'pimenta_do_reino.xlsx', 'sisal.xlsx',
448     'tomate.xlsx', 'uva.xlsx', 'cafe_total.xlsx')
449
450 #pegando item por item de agricultura pelo codigo
451 linksArquivosAgricola <- NULL
452 #o item com codigo 39466 nao existe, por isso nao eh indicado
453 for(i in 39428:39471){
454   if(i !=39466){
455     linksArquivosAgricola <- c(linksArquivosAgricola,
456       paste('name=tabela6588.xlsx&terr=N&rank=-&query=t/6588/n1/all/n3/
457         all/v/all/p/all/c48/', i, '/l/v,p%2Bc48,t', sep=""))
458   }
459 }
460 linksArquivosAgricola <- c(linksArquivosAgricola,
461   'name=tabela6588.xlsx&terr=N&rank=-&query=t/6588/n1/all/n3/
462   all/v/all/p/all/c48/40527/l/v,p%2Bc48,t')
463
464 #baixando as planilhas
465 baixarPlanilhaIBGE(nomesArquivosAgricola, linksArquivosAgricola);
466 #limpando os dados
467 dadosAgricultura <- limpaDadosAgricultura(dadosAgricultura,
468   nomesArquivosAgricola)
469
470 #removendo registros com valor 0
471 dadosAgricultura <- dadosAgricultura %>% filter(VALOR > 0)
472 dadosPecuararia <- dadosPecuararia %>% filter(VALOR > 0)
473
474 #deixando a coluna valor em numerico
475 dadosAgricultura <- mutate_each(dadosAgricultura, funs(as.numeric), VALOR)
```

```
475
476 #transformando valores da tabela de agricultura
477 dadosAgricultura <- dadosAgricultura %>%
478   mutate(VALOR=case_when(UNIDADE == "Hectares" |
479     UNIDADE == "Toneladas" ~ round(VALOR/1000, digits = 2),
480     UNIDADE != "Hectares" & UNIDADE != "Toneladas" ~ VALOR))
481 dadosAgricultura$UNIDADE[dadosAgricultura$UNIDADE == "Hectares"] <- "Mil
    Hectares"
482 dadosAgricultura$UNIDADE[dadosAgricultura$UNIDADE == "Toneladas"] <- "Mil
    Toneladas"
483 dadosAgricultura$UNIDADE[dadosAgricultura$UNIDADE == "Quilogramas por
    Hectare"] <- "Kg/Hectare"
484
485 #transformando valores da tabela de pecuaria
486 dadosPecuaria <- dadosPecuaria %>%
487   mutate(VALOR=case_when(UNIDADE == "Cabeas" ~ round(VALOR/1000, digits = 2),
488     UNIDADE == "Quilogramas" ~ round(VALOR/1000000, digits = 2),
489     UNIDADE != "Cabeas" & UNIDADE != "Quilogramas" ~ VALOR))
490 dadosPecuaria$UNIDADE[dadosPecuaria$UNIDADE == "Cabeas"] <- "Mil Cabeas"
491 dadosPecuaria$UNIDADE[dadosPecuaria$UNIDADE == "Quilogramas"] <- "Mil
    Toneladas"
492
493 }, error = function(e) {
494   print("ERRO NA EXECUCO DO IBGE!")
495   writeLines(paste0("Ocorreu um erro na limpeza dos dados do IBGE! - ", e),
496     paste0(caminhoDownloadArquivos,"erro_limpeza_IBGE_", Sys.time(), ".txt"));
497 })
498
499 ##### dados da CONAB #####
500
501 tryCatch({
502   todosTitulos = NULL
503   todosLinks = NULL
504   #percorrendo as paginas pegando titulo e link para download
505   for (i in 0:3){
506     pagina = i*10
507     url <- paste('https://www.conab.gov.br/info-agro/safras/
508       serie-historica-das-safras?limitstart=', pagina, sep="")
509     links <- read_html(url)
510
511     tituloFormatado = links %>% html_nodes(".catItemAttachments") %>%
512       html_nodes("a") %>%
```



```

512   html_attr('title')
513   todosTitulos = c(todosTitulos, tituloFormatado)
514   linksFormatado = links %>% html_nodes(".catItemAttachments") %>%
      html_nodes("a") %>%
515   html_attr('href')
516   todosLinks = c(todosLinks, linksFormatado)
517 }
518
519 #realizando download dos arquivos e salvando na pasta
520 for (j in 1:length(todosTitulos)){
521   caminhoDestino <- paste(caminhoDownloadArquivos, "CONAB/",
522                           todosTitulos[j], sep="");
523   url <- paste("https://www.conab.gov.br/", todosLinks[j], sep="");
524   download.file(url, caminhoDestino);
525 }
526
527 #entra arquivo por arquivo, deixando de lado aqueles definidos anteriormente
      que nao seria utilizados
528 for (i in 1:length(todosTitulos)){
529   if(todosTitulos[i] != "BrasilUFSerieHist.xls" &
530       todosTitulos[i] != "BrasilProdutoSerieHist.xls" &
531       todosTitulos[i] != "CanaSerieHistZ-ZIndustria.xls" &
532       todosTitulos[i] != "CanaSerieHistZ-ZreaZTotal.xls" &
533       todosTitulos[i] != "ParqueCafTotalSerieHist.xls" &
534       todosTitulos[i] != "ParqueCafTotalSerieHist.xls" &
535       todosTitulos[i] != "CafeTotalSerieHist.xls" &
536       todosTitulos[i] != "CafeConilonSerieHist.xls" &
537       todosTitulos[i] != "CafeArbicaSerieHist.xls" &
538       !grepl(".ods", todosTitulos[i])){
539     abas = excel_sheets(path = paste0("~/Downloads/CONAB/", todosTitulos[i]))
540     for(j in 1:length(abas)){
541       dados <- data.frame(1,2,3,4,5,6,7)
542       colnames(dados) <- c("ANO", "ESTADO", "FONTE", "PRODUTO", "INDICADOR",
543                           "UNIDADE", "VALOR")
544       #lendo excel
545       produto <- read_excel(paste0("~/Downloads/CONAB/", todosTitulos[i]),
546                             sheet = abas[j])
547       #pegando nome do produto, contido no cabealho
548       nomeProduto = unlist(strsplit(colnames(produto[1]), " - "))[1]
549       linhaAPegarNomeIndicador = 1
550       if(nomeProduto == "X__1"){
551         nomeProduto = unlist(strsplit(produto$X__1[1], " - "))[1]

```

```
550     linhaAPegarNomeIndicador = 2
551   }
552   #renomeando cabealho
553   colnames(produto) <- c("estado");
554   #pegando nome do indicador, logo na primeira ou segunda linha
555   nomeIndicador =
       stringr::str_replace_all(produto$estado[linhaAPegarNomeIndicador],
       "Srie Histrica de ", "")
556   if(linhaAPegarNomeIndicador == 1){
557     produto <- produto[-c(1, 2), ]
558   }else{
559     produto <- produto[-c(1, 2, 3), ]
560   }
561
562   unidade = stringr::str_replace_all(produto$estado[1], "Em ", "")
563
564   #tornando a segunda linha como cabealho
565   colnames(produto) <- c(produto[2,])
566
567   produto <- produto[-c(1, 2), ]
568
569   #comecando a inserir na planilha
570   totalLinhas = nrow(produto)
571   totalColunas = ncol(produto)
572
573   for(linha in 1:totalLinhas){
574     for(coluna in 2:totalColunas){
575       ano = substr(colnames(produto[coluna]), 1, 4)
576       valor = as.double(produto[linha,coluna])
577       if(!is.na(valor) & valor > 0){
578         dados <- dados %>%
579           add_row(ANO = ano,
580                 ESTADO = produto$'REGIO/UF'[linha],
581                 FONTE = "CONAB",
582                 PRODUTO = nomeProduto,
583                 INDICADOR = nomeIndicador,
584                 UNIDADE = unidade,
585                 VALOR = valor)
586       }
587     }
588   }
589   dadosAgricultura <- rbind(dadosAgricultura, dados)
```

```
590     }
591   }
592 }
593
594 #limpeza nos dados
595
596 #removendo dados com valor 0 ou null
597 dadosAgricultura <- dadosAgricultura %>% filter(VALOR > 0)
598 dadosAgricultura <- dadosAgricultura %>% filter(ESTADO != "2")
599
600 #limpeza de dados que saíram com dados diferentes dos padroes
601 dadosAgricultura$INDICADOR[dadosAgricultura$INDICADOR == "Produto de
    Cana-de-Acar"] <- " Produto "
602 dadosAgricultura$INDICADOR[dadosAgricultura$INDICADOR == "Srie Histrica da rea
    de Plantio"] <- "rea Plantada"
603
604 #removendo registros das regies centro-sul e norte/nordeste
605 dadosAgricultura <- dadosAgricultura %>%
606   filter(ESTADO != "CENTRO-SUL" & ESTADO != "NORTE/NORDESTE")
607
608 }, error = function(e) {
609   print("ERRO NA EXECUCO DA CONAB")
610   writeLines(paste0("Ocorreu um erro na limpeza dos dados da CONAB! - ",
    todosTitulos[i], " - ", e),
    paste0(caminhoDownloadArquivos,"erro_limpeza_CONAB_", Sys.time(),
    ".txt"));
611 })
612
613 ##### dados da EMBRAPA #####
614
615 tryCatch({
616   arquivos = list.files(paste0(caminhoDownloadArquivos, "EMBRAPA/Agricultura"))
617   arquivosPecuarria = list.files(paste0(caminhoDownloadArquivos,
    "EMBRAPA/Pecuarria"))
618
619 #funcao para gerar registros de totais por regioao
620 geraRegistrosRegiaoEmbrapa <- function(dados, nomeIndicador, nomeProduto,
    nomeUnidade) {
621   anos <- unique(dados$ANO)
622   dados <- mutate_each(dados, funs(as.numeric), VALOR)
623   for(i in 1:length(anos)){
624     #fazendo a soma do brasil
```

```
625     registrosSul = nrow(dados %>%
626       filter(ANO == anos[i] &
627         PRODUTO == nomeProduto &
628         INDICADOR == nomeIndicador &
629         FONTE == "EMBRAPA" &
630         (ESTADO == "PR" | ESTADO == "Paran" |
631         ESTADO == "RS" | ESTADO == "Rio Grande do Sul" |
632         ESTADO == "SC" | ESTADO == "Santa Catarina")) %>%
633       select(VALOR))
634   if(registrosSul > 0){
635     totalSul = sum(dados %>%
636       filter(ANO == anos[i] &
637         PRODUTO == nomeProduto &
638         INDICADOR == nomeIndicador &
639         FONTE == "EMBRAPA" &
640         (ESTADO == "PR" | ESTADO == "Paran" |
641         ESTADO == "RS" | ESTADO == "Rio Grande do Sul" |
642         ESTADO == "SC" | ESTADO == "Santa Catarina")) %>%
643       select(VALOR))
644     dados[nrow(dados)+1, ] <- data.frame(ANO = anos[i],
645       ESTADO = "Sul",
646       FONTE = "EMBRAPA",
647       PRODUTO = nomeProduto,
648       INDICADOR = nomeIndicador,
649       UNIDADE = nomeUnidade,
650       VALOR = totalSul,
651       stringsAsFactors = FALSE)
652   }
653   #fazendo somatorio dos dados da regioao norte
654   registrosNorte = nrow(dados %>%
655     filter(ANO == anos[i] &
656       PRODUTO == nomeProduto &
657       INDICADOR == nomeIndicador &
658       FONTE == "EMBRAPA" &
659       (ESTADO == "RO" | ESTADO == "Rondonia" |
660       ESTADO == "AC" | ESTADO == "Acre" |
661       ESTADO == "AM" | ESTADO == "Amazonas" |
662       ESTADO == "PA" | ESTADO == "Par" |
663       ESTADO == "RR" | ESTADO == "Roraima" |
664       ESTADO == "AP" | ESTADO == "Amap" |
665       ESTADO == "TO" | ESTADO == "Tocantins")) %>%
666     select(VALOR))
```

```
667     if(registrosNorte > 0){
668         totalNorte = sum(dados %>%
669             filter(ANO == anos[i] &
670                 PRODUTO == nomeProduto &
671                 INDICADOR == nomeIndicador &
672                 FONTE == "EMBRAPA" &
673                 (ESTADO == "RO" | ESTADO == "Rondnia" |
674                 ESTADO == "AC" | ESTADO == "Acre" |
675                 ESTADO == "AM" | ESTADO == "Amazonas" |
676                 ESTADO == "PA" | ESTADO == "Par" |
677                 ESTADO == "RR" | ESTADO == "Roraima" |
678                 ESTADO == "AP" | ESTADO == "Amap" |
679                 ESTADO == "TO" | ESTADO == "Tocantins"))) %>%
680             select(VALOR))
681         dados[nrow(dados)+1, ] <- data.frame(ANO = anos[i],
682                                             ESTADO = "Norte",
683                                             FONTE = "EMBRAPA",
684                                             PRODUTO = nomeProduto,
685                                             INDICADOR = nomeIndicador,
686                                             UNIDADE = nomeUnidade,
687                                             VALOR = totalNorte,
688                                             stringsAsFactors = FALSE)
689     }
690     #fazendo somatorio dos dados da regioao nordeste
691     registrosNordeste = nrow(dados %>%
692         filter(ANO == anos[i] &
693             PRODUTO == nomeProduto &
694             INDICADOR == nomeIndicador &
695             FONTE == "EMBRAPA" &
696             (ESTADO == "MA" | ESTADO == "Maranhao" |
697             ESTADO == "PI" | ESTADO == "Piau" |
698             ESTADO == "CE" | ESTADO == "Cear" |
699             ESTADO == "BA" | ESTADO == "BAHIA" |
700             ESTADO == "RN" | ESTADO == "Rio Grande do Norte" |
701             ESTADO == "PB" | ESTADO == "Paraba" |
702             ESTADO == "PE" | ESTADO == "Pernambuco" |
703             ESTADO == "AL" | ESTADO == "Alagoas" |
704             ESTADO == "SE" | ESTADO == "Sergipe"))) %>%
705         select(VALOR))
706     if(registrosNordeste > 0){
707         totalNordeste = sum(dados %>%
708             filter(ANO == anos[i] &
```

```

709         PRODUTO == nomeProduto &
710         INDICADOR == nomeIndicador &
711         FONTE == "EMBRAPA" &
712         (ESTADO == "MA" | ESTADO == "Maranhão" |
713         ESTADO == "PI" | ESTADO == "Piau" |
714         ESTADO == "CE" | ESTADO == "Cear" |
715         ESTADO == "BA" | ESTADO == "BAHIA" |
716         ESTADO == "RN" | ESTADO == "Rio Grande do Norte" |
717         ESTADO == "PB" | ESTADO == "Paraba" |
718         ESTADO == "PE" | ESTADO == "Pernambuco" |
719         ESTADO == "AL" | ESTADO == "Alagoas" |
720         ESTADO == "SE" | ESTADO == "Sergipe")) %>%
721         select(VALOR))
722     dados[nrow(dados)+1, ] <- data.frame(ANO = anos[i],
723                                           ESTADO = "Nordeste",
724                                           FONTE = "EMBRAPA",
725                                           PRODUTO = nomeProduto,
726                                           INDICADOR = nomeIndicador,
727                                           UNIDADE = nomeUnidade,
728                                           VALOR = totalNordeste,
729                                           stringsAsFactors = FALSE)
730 }
731 #fazendo somatorio dos dados da regioao centro-oeste
732 registrosCentroOeste = nrow(dados %>%
733     filter(ANO == anos[i] &
734           PRODUTO == nomeProduto &
735           INDICADOR == nomeIndicador &
736           FONTE == "EMBRAPA" &
737           (ESTADO == "MT" | ESTADO == "Mato Grosso" |
738           ESTADO == "MS" | ESTADO == "Mato Grosso do Sul" |
739           ESTADO == "GO" | ESTADO == "Gois" |
740           ESTADO == "DF" | ESTADO == "Distrito Federal")) %>%
741           select(VALOR))
742 if(registrosCentroOeste > 0){
743     totalCentroOeste = sum(dados %>%
744         filter(ANO == anos[i] &
745               PRODUTO == nomeProduto &
746               INDICADOR == nomeIndicador &
747               FONTE == "EMBRAPA" &
748               (ESTADO == "MT" | ESTADO == "Mato Grosso" |
749               ESTADO == "MS" | ESTADO == "Mato Grosso do Sul" |
750               ESTADO == "GO" | ESTADO == "Gois" |

```



```
793   }
794   return(dados)
795 }
796
797 for(a in 1:length(arquivos)){
798   produto <- read_excel(paste0(caminhoDownloadArquivos,"EMBRAPA/Agricultura/",
799     arquivos[a]))
800
801   dados <- data.frame(1,2,3,4,5,6,7)
802   colnames(dados) <- c("ANO", "ESTADO", "FONTE", "PRODUTO","INDICADOR",
803     "UNIDADE", "VALOR")
804
805   #removendo colunas no utilizadas
806   produto <- select(produto, -starts_with("Rendimento"))
807   produto <- select(produto, -starts_with("Valor"))
808
809   #renomeando colunas
810   names(produto) <- c("ANO", "UF", "rea Colhida", "Produo")
811
812   #pegando o nome do produto
813   nomeProduto = gsub(".xlsx", "", arquivos[a])
814   nomeProduto = gsub("_", " ", nomeProduto)
815
816   auxiliarAno = 0
817   totalRegistros = 1
818   for(i in 1:nrow(produto)){
819     ano = produto$ANO[i]
820     if(is.na(ano) | ano == 0){
821       ano = auxiliarAno
822     }else{
823       auxiliarAno = ano
824     }
825
826     UF = produto$UF[i]
827     if(UF == "Total:"){
828       UF = "Brasil"
829     }
830
831     #adicionando na tabela
832     if(!is.na(produto$`rea Colhida`[i])){
833       produtividade = (produto$Produo[i]/produto$`rea Colhida`[i]) * 1000
834       #inserindo rea colhida
```

```
833     dados <- dados %>%
834       add_row(ANO = ano,ESTADO = UF,
835             FONTE = "EMBRAPA",
836             PRODUTO = nomeProduto,
837             INDICADOR ="rea Colhida",
838             UNIDADE = "Mil Hectares",
839             VALOR = sprintf("%.3f",produto$`rea Colhida`[i]))
840   #inserindo produo
841   dados <- dados %>%
842     add_row(ANO = ano,
843           ESTADO = UF,
844           FONTE = "EMBRAPA",
845           PRODUTO = nomeProduto,
846           INDICADOR =" Produo ",
847           UNIDADE = "Mil Toneladas",
848           VALOR = sprintf("%.3f",produto$Produo[i]))
849   #inserindo produtividade
850   dados <- dados %>%
851     add_row(ANO = ano,
852           ESTADO = UF,
853           FONTE = "EMBRAPA",
854           PRODUTO = nomeProduto,
855           INDICADOR ="Produtividade",
856           UNIDADE = "Kg/Hectare",
857           VALOR = sprintf("%.3f",produtividade))
858   totalRegistros = totalRegistros+3
859 }
860 }
861 dados = geraRegistrosRegiaoEmbrapa(dados, "Produo", nomeProduto, "Mil
      Toneladas")
862 dados = geraRegistrosRegiaoEmbrapa(dados, "Produtividade", nomeProduto,
      "Kg/Hectare")
863 dados = geraRegistrosRegiaoEmbrapa(dados, "rea Colhida", nomeProduto, "Mil
      Hectares")
864 dadosAgricultura <- rbind(dadosAgricultura, dados)
865 }
866
867 for(a in 1:length(arquivosPecuararia)){
868   produto <- read_excel(paste0(caminhoDownloadArquivos, "EMBRAPA/Pecuararia/",
      arquivosPecuararia[a]))
869   dados <- data.frame(1,2,3,4,5,6,7)
```

```

870   colnames(dados) <- c("ANO", "ESTADO", "FONTE", "PRODUTO","INDICADOR",
      "UNIDADE", "VALOR")
871   if(arquivosPecuaria[a] == "Aquicultura.xlsx" | arquivosPecuaria[a] ==
      "Produtos_Origem_Animal.xlsx" | arquivosPecuaria[a] == "Rebanhos.xlsx"){
872
873     produto <- select(produto, -starts_with("Valor"))
874     produto <- select(produto, -starts_with("Preo"))
875
876     indicador = " Produto "
877     unidade = "Mil Toneladas"
878     if(arquivosPecuaria[a] == "Rebanhos.xlsx"){
879       indicador = "Rebanho"
880       unidade = "Mil Cabeas"
881     }
882
883     #renomeando colunas
884     names(produto) <- c("ANO", "UF", "Produto", "Produto")
885     totalRegistros = 1
886     for(i in 1:nrow(produto)){
887       #adicionando na tabela
888       if(!is.na(produto$Produto [i]) & produto$Produto [i] > 0){
889         valor = produto$Produto [i]/1000000
890         nomeProduto = gsub(" \\(Quilogramas\\)", "", produto$Produto[i])
891         nomeProduto = gsub(" \\(Milheiros\\)", "", nomeProduto)
892         nomeProduto = gsub(" \\(Nenhuma\\)", "", nomeProduto)
893         nomeProduto = gsub(" \\(Mil dzias\\)", "", nomeProduto)
894         nomeProduto = gsub(" \\(Mil litros\\)", "", nomeProduto)
895         if(indicador == "Rebanho"){
896           valor = produto$Produto [i]/1000;
897         }
898         #inserindo registro
899         dados <- dados %>%
900           add_row(ANO = produto$ANO[i],
901                 ESTADO = produto$UF[i],
902                 FONTE = "EMBRAPA",
903                 PRODUTO = nomeProduto,
904                 INDICADOR = indicador,
905                 UNIDADE = unidade,
906                 VALOR = sprintf("%.3f",valor))
907         totalRegistros = totalRegistros+1
908       }
909     }

```

```
910     nomeProdutos = unique(produto$Produto)
911     for(i in 1:length(nomeProdutos)){
912         if(arquivosPecuaria[a] != "Rebanhos.xlsx"){
913             dados = geraRegistrosRegiaoEmbrapa(dados, "Produo", nomeProdutos[i],
914                 "Mil Toneladas")
915         }else{
916             dados = geraRegistrosRegiaoEmbrapa(dados, "Rebanho", nomeProdutos[i],
917                 "Mil Cabeas")
918         }
919     }
920     dadosPecuaria <- rbind(dadosPecuaria, dados)
921 }else if(arquivosPecuaria[a] == "Abate.xlsx"){
922     #removendo coluna tipo de inspecao e total informantes
923     produto <- select(produto, -starts_with("Tipo de Inspeo"))
924     produto <- select(produto, -starts_with("Nmero de Informantes"))
925     names(produto) <- c("ANO", "UF", "Produto", "Rebanho", "Produo")
926     for(i in 1:nrow(produto)){
927         valor = produto$Rebanho[i] / 1000;
928         if(valor > 0){
929             dados <- dados %>%
930                 add_row(ANO = produto$ANO[i],
931                     ESTADO = produto$UF[i],
932                     FONTE = "EMBRAPA",
933                     PRODUTO = paste0("Abate - ", produto$Produto[i]),
934                     INDICADOR = "Rebanho",
935                     UNIDADE = "Mil Cabeas",
936                     VALOR = sprintf("%.3f",valor))
937         }
938     }
939     valor = produto$Produo [i] / 1000;
940     if(valor > 0){
941         dados <- dados %>%
942             add_row(ANO = produto$ANO[i],
943                 ESTADO = produto$UF[i],
944                 FONTE = "EMBRAPA",
945                 PRODUTO = paste0("Abate - ", produto$Produto[i]),
946                 INDICADOR = "Produo ",
947                 UNIDADE = "Mil Toneladas",
948                 VALOR = sprintf("%.3f",valor))
949     }
950 }
951 nomeProdutos = unique(produto$Produto)
952 for(i in 1:length(nomeProdutos)){
```

```
950     dados = geraRegistrosRegiaoEmbrapa(dados, "Produo", paste0("Abate - ",
951       nomeProdutos[i]), "Mil Toneladas")
952   }
953   dadosPecuaria <- rbind(dadosPecuaria, dados)
954 }else if(arquivosPecuaria[a] == "Ovinos_Tosqueados.xlsx" |
955   arquivosPecuaria[a] == "Vacas_Ordenhadas.xlsx"){
956   nomeProduto = "Ovinos Tosqueados"
957   nomeProduto2 = "L"
958   unidade = "Kg"
959   unidade2 = "Kg/Animal"
960   if(arquivosPecuaria[a] == "Vacas_Ordenhadas.xlsx"){
961     nomeProduto = "Vacas Leiteiras"
962     nomeProduto2 = "Leite"
963     unidade = "Mil Litros"
964     unidade2 = "Litros/Animal"
965   }
966   #renomeando colunas
967   nomes(produto) <- c("ANO", "UF", "Rebanho", "Produo", "Produtividade")
968   for(i in 1:nrow(produto)){
969     #adicionando na tabela
970     if(!is.na(produto$Rebanho[i]) & produto$Rebanho[i] > 0){
971       #inserindo registro de rebanho
972       valor = produto$Rebanho[i]/1000
973       dados <- dados %>%
974         add_row(ANO = produto$ANO[i],
975           ESTADO = produto$UF[i],
976           FONTE = "EMBRAPA",
977           PRODUTO = nomeProduto,
978           INDICADOR = "Rebanho",
979           UNIDADE = "Mil Cabeas",
980           VALOR = sprintf("%.3f",valor))
981     }
982     if(!is.na(produto$Produo [i]) & produto$Produo > 0){
983       #inserindo registro de rebanho
984       dados <- dados %>%
985         add_row(ANO = produto$ANO[i],
986           ESTADO = produto$UF[i],
987           FONTE = "EMBRAPA",
988           PRODUTO = nomeProduto2,
```

```
989         INDICADOR = " Produto ",
990         UNIDADE = unidade,
991         VALOR = sprintf("%.3f",produto$Produto[i]))
992     }
993     if(!is.na(produto$Produtividade[i]) & produto$Produtividade > 0){
994         #inserindo registro de rebanho
995         dados <- dados %>%
996             add_row(ANO = produto$ANO[i],
997                 ESTADO = produto$UF[i],
998                 FONTE = "EMBRAPA",
999                 PRODUTO = nomeProduto2,
1000                 INDICADOR = "Produtividade",
1001                 UNIDADE = unidade2,
1002                 VALOR = sprintf("%.3f",produto$Produtividade[i]))
1003     }
1004 }
1005 dados = geraRegistrosRegiaoEmbrapa(dados, "Rebanho", nomeProduto, "Mil
1006                                     Cabeas")
1007 dados = geraRegistrosRegiaoEmbrapa(dados, "Produto", nomeProduto2, unidade)
1008 dados = geraRegistrosRegiaoEmbrapa(dados, "Produtividade",
1009                                     nomeProdutos[i], unidade2)
1010 dadosPecuararia <- rbind(dadosPecuararia, dados)
1011 }
1012 }, error = function(e) {
1013     print("ERRO NA EXECUCO DA EMBRAPA")
1014     writeLines(paste0("Ocorreu um erro na limpeza dos dados da EMBRAPA! - ", e),
1015               paste0(caminhoDownloadArquivos,"erro_limpeza_EMBRAPA_", Sys.time(),
1016                     ".txt"));
1017 }
1018 }
1019 ##### dados da FAO #####
1020
1021 tryCatch({
1022     anoAtual = format(Sys.Date(), "%Y")
1023     ano = 1961
1024     linkExtraido = 1
1025     stringAnos = "1961"
1026     #montando a string de anos para a pesquisa, de 1961 at o ano atual, separados
1027     por vrgula
1028     while(ano<anoAtual){
1029         ano = ano+1
```

```

1026   stringAnos = paste(stringAnos, ",", ano, sep="")
1027 }
1028
1029 #o primeiro link contem os dados de agricultura, o segundo de animais vivos, o
      terceiro de produtos de origem animal
1030 link <-
1031   c(paste("http://fenixservices.fao.org/faostat/api/v1/en/data/QC?area=21&
1032   area_cs=FA0&element=2312,2510,2413&item=800,221,711,515,526,226,366,367,572,
1033   203,486,44,782,176,414,558,552,216,181,89,358,101,461,426,217,591,125,378,265,
1034   393,108,531,530,220,191,459,689,401,693,698,661,249,656,813,767,329,195,554,
1035   397,550,577,399,821,569,773,94,512,619,542,541,603,406,720,549,103,507,560,242,
1036   839,225,777,336,677,277,780,778,310,311,263,592,224,407,497,201,372,333,210,56,
1037   446,571,809,671,568,299,79,449,292,702,234,75,254,257,339,430,260,403,402,490,
1038   256,600,534,521,187,417,687,748,587,197,574,223,489,536,296,116,211,394,754,523,
1039   92,788,270,547,27,149,836,71,280,328,289,789,83,236,723,373,544,423,157,156,161,
1040   267,122,305,495,136,667,826,388,97,275,692,463,420,205,222,567,15,137,135&
1041   item_cs=FA0&year=", stringAnos,"&show_codes=true&show_unit=true&show_flags=
1042   true&null_values=false&page_number=", sep=""),
1043   paste("http://fenixservices.fao.org/faostat/api/v1/en/data/QA?area=21&
1044   area_cs=FA0&element=2111&item=1171,1107,1181,946,1157,1126,866,1057,1068,1072,
1045   1016,1096,1110,1083,1034,1140,1150,976,1079&item_cs=FA0&year=", stringAnos,
1046   "&show_codes=true&show_unit=true&show_flags=true&null_values=false&
1047   page_number=", sep=""),
1048   paste("http://fenixservices.fao.org/faostat/api/v1/en/data/QL?area=21&
1049   area_cs=FA0&element=2313,2510,2413&item=1183,1062,1067,1091,1092,1100,957,
1050   919,1182,1122,1084,972,1137,944,1094,1070,1077,1032,1120,1124,1161,1055,1144,
1051   1154,1012,1087,1108,1089,947,1127,867,1058,1069,1163,1017,1073,1097,1111,1166,
1052   1158,1151,1035,1141,977,1080,951,1130,882,1020,982,1167,1083,1185,1195,1025,995,
1053   999,1176,987&item_cs=FA0&year=", stringAnos,
      "&show_codes=true&show_unit=true&
1054   show_flags=true&null_values=false&page_number=", sep=""))
1055
1056
1057 #chamando os links enquanto houverem dados na pagina
1058 for (site in 1:length(link)){
1059   haDados = TRUE;
1060   pagina = 1;
1061   while(haDados){
1062     json_file <- paste(link[site], pagina
      ,"&show_codes=true&show_unit=true&show_flags=true&null_values=false&
1063     page_number=", pagina, "&page_size=100&output_type=objects", sep="");
1064     json_data <- fromJSON(paste(readLines(json_file), collapse=""))

```



```
1065     pagina = pagina+1
1066     if(length(json_data$data) < 100 || length(json_data$data) == 0){
1067         haDados = FALSE;
1068     }
1069     #os dados do JSON ficam dentro da varivel json_data$data, prontos para
        serem limpos e transformados
1070     if(linkExtraido == 1){
1071         for (i in 1:length(json_data$data)){
1072             ano      = json_data$data[[i]]$'Year Code'
1073             produto = json_data$data[[i]]$Item
1074             indicador = json_data$data[[i]]$Element
1075             unidade = json_data$data[[i]]$Unit
1076             valor = json_data$data[[i]]$Value
1077             if(is.null(valor)){
1078                 valor = 0
1079             }
1080             if(unidade == "ha"){
1081                 unidade = "Mil Hectares"
1082                 valor = valor / 1000
1083             }else if(unidade == "hg/ha"){
1084                 unidade = "Kg/Hectare"
1085                 valor = valor / 10
1086             }else if(unidade == "tonnes"){
1087                 unidade = "Mil Toneladas"
1088                 valor = valor / 1000
1089             }
1090
1091             dadosAgricultura[nrow(dadosAgricultura)+1, ] <- data.frame(ANO = ano,
1092                 ESTADO = "Brasil",
1093                 FONTE = "FAO",
1094                 PRODUTO = produto,
1095                 INDICADOR = indicador,
1096                 UNIDADE = unidade,
1097                 VALOR = valor,
1098                 stringsAsFactors = FALSE)
1099         }
1100     }else{
1101         for (i in 1:length(json_data$data)){
1102             ano      = json_data$data[[i]]$'Year Code'
1103             produto = json_data$data[[i]]$Item
1104             indicador = json_data$data[[i]]$Element
1105             unidade = json_data$data[[i]]$Unit
```

```
1106     valor = json_data$data[[i]]$Value
1107     if(is.null(valor)){
1108         valor = 0
1109     }
1110     if(unidade == "tonnes"){
1111         unidade = "Mil Toneladas"
1112         valor = valor / 1000
1113     }else if(unidade == "1000 Head"){
1114         unidade = "Mil Cabeas"
1115     }else if(unidade == "Head"){
1116         unidade = "Mil Cabeas"
1117         valor = valor / 1000
1118     }else if(unidade == "100mg/An"){
1119         unidade = "Kg/Animal"
1120         valor = valor / 1000000
1121     }else if(unidade == "hg/An"){
1122         unidade = "Kg/Animal"
1123         valor = valor / 10
1124     }else if(unidade == "hg"){
1125         unidade = "Kg/Animal"
1126         valor = valor / 10
1127     }else if(unidade == "0.1g/An"){
1128         unidade = "Kg/Animal"
1129         valor = valor / 10000
1130     }else if(unidade == "1000 No"){
1131         unidade = "Mil unidades"
1132     }else if(unidade == "No"){
1133         unidade = "Mil unidades"
1134         valor = valor / 1000
1135     }
1136     dadosPecuaria[nrow(dadosPecuaria)+1, ] <- data.frame(ANO = ano,
1137         ESTADO = "BR",
1138         FONTE = "FAO",
1139         PRODUTO = produto,
1140         INDICADOR = indicador,
1141         UNIDADE = unidade,
1142         VALOR = valor,
1143         stringsAsFactors = FALSE)
1144     }
1145 }
1146 }
1147 linkExtraido = linkExtraido+1
```

```
1148 }
1149
1150 #traduzindo nome dos produtos de agricultura
1151 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Agave fibres nes"] <-
    "Fibra Agave"
1152 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Almonds, with shell"] <-
    "Amndoas com Casca"
1153 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Anise, badian, fennel,
    coriander"] <- "Anis, Badian, Erva-Doce, Coentro"
1154 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Apples"] <- "Ma"
1155 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Apricots"] <- "Damasco"
1156 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Areca nuts"] <- "Noz de
    Areca"
1157 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Artichokes"] <-
    "Alcachofra"
1158 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Asparagus"] <- "Aspargo"
1159 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Avocados"] <- "Abacates"
1160 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Bambara beans"] <-
    "Feijo de bambara"
1161 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Bananas"] <- "Banana"
1162 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Barley"] <- "Cevada"
1163 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Bastfibres, other"] <-
    "Fibra de Bast"
1164 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Beans, dry"] <- "Feijo
    seco"
1165 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Beans, green"] <- "Feijo
    Verde"
1166 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Berries nes"] <- "Frutas
    Silvestres"
1167 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Blueberries"] <- "Amora"
1168 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Brazil nuts, with
    shell"] <- "Castanha-do-Par"
1169 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Broad beans, horse
    beans, dry"] <- "Feijo Cavalo"
1170 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Buckwheat"] <- "Trigo
    Mourisco"
1171 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Cabbages and other
    brassicas"] <- "Couve"
1172 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Canary seed"] <-
    "Alpiste"
1173 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Carobs"] <- "Alfarroba"
```

```
1174 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Carrots and turnips"] <-  
      "Cenoura e Nabo"  
1175 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Cashew nuts, with  
      shell"] <- "Castanha de Caju"  
1176 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Cashewapple"] <- "Caju"  
1177 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Cassava"] <- "Mandioca"  
1178 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Cassava leaves"] <-  
      "Folhas de mandioca"  
1179 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Castor oil seed"] <-  
      "Semente de mamona"  
1180 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Cauliflowers and  
      broccoli"] <- "Couve-flor e brcolis"  
1181 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Cereals, nes"] <-  
      "Cereais"  
1182 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Cherries"] <- "Cereja"  
1183 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Cherries, sour"] <-  
      "Cereja Azeda"  
1184 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Chestnut"] <- "Castanha"  
1185 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Chick peas"] <- "Gro de  
      Bico"  
1186 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Chicory roots"] <- "Raz  
      de Chicria"  
1187 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Chillies and peppers,  
      dry"] <- "Pimenta"  
1188 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Chillies and peppers,  
      green"] <- "Pimenta Verde"  
1189 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Cinnamon (canella)"] <-  
      "Canela"  
1190 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Cloves"] <- "Cravo"  
1191 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Cocoa, beans"] <-  
      "Sementes de cacau"  
1192 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Coconuts"] <- "Cco"  
1193 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Coffee, green"] <- "Caf  
      Verde"  
1194 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Coir"] <- "Fibra de Cco"  
1195 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Cotton lint"] <- "Fibra  
      de Algodó"  
1196 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Cottonseed"] <- "Algodó"  
1197 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Cow peas, dry"] <-  
      "Ervilha"  
1198 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Cranberries"] <-  
      "Oxicoco"
```

```

1199 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Cucumbers and gherkins"]
      <- "Pepino"
1200 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Currants"] <- "Groselha"
1201 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Dates"] <- "Tmara"
1202 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Eggplants (aubergines)"]
      <- "Beringela"
1203 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Fibre crops nes"] <-
      "Fibras"
1204 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Figs"] <- "Figo"
1205 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Flax fibre and tow"] <-
      "Fibra de Linho"
1206 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Fonio"] <- "Fonio"
1207 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Fruit, citrus nes"] <-
      "Frutas Ctricas"
1208 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Fruit, fresh nes"] <-
      "Frutas Frescas"
1209 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Fruit, pome nes"] <-
      "Frutas de Pomar"
1210 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Fruit, stone nes"] <-
      "Drupas"
1211 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Fruit, tropical fresh
      nes"] <- "Frutas Tropicais"
1212 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Garlic"] <- "Alho"
1213 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Ginger"] <- "Gengibre"
1214 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Gooseberries"] <-
      "Groselha"
1215 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Grain, mixed" ] <- "Gros"
1216 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Grapefruit (inc.
      pomelos)"] <- "Toranja"
1217 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Grapes"] <- "Uva"
1218 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Groundnuts, with shell"]
      <- "Amendoim com Casca"
1219 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Gums, natural"] <- "Goma"
1220 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Hazelnuts, with shell"]
      <- "Avel com Casca"
1221 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Hemp tow waste"] <-
      "Resduo de Cnhamo"
1222 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Hempseed"] <- "Semente
      de Cnhamo"
1223 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Hops"] <- "Ipulo"
1224 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Jojoba seed"] <- "Jojoba"
1225 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Jute"] <- "Juta"

```

```
1226 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Kapok fibre"] <- "Fibra
    de Samama"
1227 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Kapok fruit" ] <-
    "Samama"
1228 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Kapokseed in shell"] <-
    "Semente de Samama"
1229 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Karite nuts (sheanuts)"]
    <- "Noz de Carit"
1230 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Kiwi fruit"] <- "Kiwi"
1231 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Kola nuts"] <-
    "Noz-de-Cola"
1232 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Leeks, other alliaceous
    vegetables"] <- "Alho-Por"
1233 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Lemons and limes"] <-
    "Limo"
1234 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Lentils"] <- "Lentilha"
1235 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Lettuce and chicory"] <-
    "Alface e Chicria"
1236 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Linseed"] <- "Linhaa"
1237 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Lupins"] <- "Tremoo"
1238 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Maize"] <- "Milho"
1239 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Maize, green"] <- "Milho
    Verde"
1240 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Mangoes, mangosteens,
    guavas"] <- "Manga, Mangosto, Goiaba"
1241 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Manila fibre (abaca)"]
    <- "Cnhamo de Manila"
1242 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Mat"] <- "Mate"
1243 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Melons, other
    (inc.cantaloupes)"] <- "Melo"
1244 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Melonseed"] <- "Semente
    de Melo"
1245 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Millet"] <- "Paino"
1246 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Mushrooms and truffles"]
    <- "Cogumelo e Trufa"
1247 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Mustard seed"] <-
    "Semente de Mostarda"
1248 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Nutmeg, mace and
    cardamoms"] <- "Noz-Moscada e Cardamomo"
1249 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Nuts, nes"] <- "Noz"
1250 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Oats"] <- "Aveia"
```

```
1251 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Oil palm fruit"] <-  
      "Fruta de leo de Palma"  
1252 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Oil, palm"] <- "leo de  
      Palma"  
1253 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Oilseeds nes"] <-  
      "Oleaginosas"  
1254 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Okra"] <- "Quiabo"  
1255 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Olives"] <- "Azeitona"  
1256 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Onions, dry"] <- "Cebola"  
1257 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Onions, shallots,  
      green"] <- "Cebolinha"  
1258 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Oranges"] <- "Laranja"  
1259 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Palm kernels"] <- "Gro  
      de Palma"  
1260 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Papayas"] <- "Mamo"  
1261 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Peaches and nectarines"]  
      <- "Pssego e Nectarina"  
1262 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Pears"] <- "Pra"  
1263 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Peas, dry"] <- "Ervilha  
      Seca"  
1264 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Peas, green"] <-  
      "Ervilha Verde"  
1265 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Pepper (piper spp.)"] <-  
      "Pimenta"  
1266 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Peppermint"] <- "Hortel"  
1267 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Persimmons"] <- "Caqui"  
1268 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Pigeon peas"] <- "Guandu"  
1269 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Pineapples"] <- "Abacaxi"  
1270 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Pistachios"] <-  
      "Pistache"  
1271 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Plantains and others"]  
      <- "Pltano"  
1272 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Plums and sloes"] <-  
      "Ameixa"  
1273 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Poppy seed"] <- "Semente  
      de Papoula"  
1274 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Potatoes"] <- "Batata"  
1275 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Pulses, nes"] <-  
      "Legumes"  
1276 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Pumpkins, squash and  
      gourds"] <- "Abbora e Cabaa"
```



```
1277 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Pyrethrum, dried"] <-  
      "Pretro"  
1278 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Quinces"] <- "Marmelo"  
1279 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Quinoa"] <- "Quinoa"  
1280 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Ramie"] <- "Rami"  
1281 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Rapeseed"] <- "Colza"  
1282 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Raspberries"] <-  
      "Framboesa"  
1283 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Rice, paddy"] <- "Arroz"  
1284 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Roots and tubers, nes"]  
      <- "Razes e Tubrculos"  
1285 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Rubber, natural"] <-  
      "Borracha"  
1286 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Rye"] <- "Centeio"  
1287 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Safflower seed"] <-  
      "Semente de Crtamo"  
1288 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Seed cotton" ] <-  
      "Semente de Algodó"  
1289 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Sesame seed"] <-  
      "Semente de Gergilim"  
1290 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Sisal"] <- "Sisal"  
1291 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Sorghum"] <- "Sorgo"  
1292 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Soybeans"] <- "Soja"  
1293 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Spices, nes"] <-  
      "Especiarias"  
1294 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Spinach"] <- "Espinafre"  
1295 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Strawberries"] <-  
      "Morango"  
1296 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "String beans"] <- "Vagem"  
1297 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Sugar beet"] <-  
      "Beterraba"  
1298 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Sugar cane"] <-  
      "Cana-de-Acar"  
1299 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Sugar crops, nes"] <-  
      "Cana-de-Acar"  
1300 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Sunflower seed" ] <-  
      "Semente de Girassol"  
1301 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Sweet potatoes"] <-  
      "Batata Doce"  
1302 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Tallowtree seed"] <-  
      "Semente de Sebo"
```



```
1303 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Tangerines, mandarins,
      clementines, satsumas"] <- "Tangerina, Clementina e Satsumo"
1304 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Taro (cocoyam)"] <-
      "Inhame"
1305 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Tea"] <- "Ch"
1306 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Tobacco,
      unmanufactured"] <- "Tabaco"
1307 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Tomatoes"] <- "Tomate"
1308 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Triticale"] <-
      "Triticale"
1309 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Tung nuts"] <- "Noz de
      Tungue"
1310 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Vanilla"] <- "Baunilha"
1311 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Vegetables, fresh nes"]
      <- "Legumes Frescos"
1312 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Vegetables, leguminous
      nes"] <- "Leguminosas"
1313 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Vetches"] <- "Vicia"
1314 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Walnuts, with shell"] <-
      "Noz com Casca"
1315 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Watermelons"] <-
      "Melancia"
1316 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Wheat"] <- "Trigo"
1317 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Yams"] <- "Inhame"
1318 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Yautia (cocoyam)"] <-
      "Taioba"
1319
1320 #traduzindo o nome dos indicadores
1321 dadosAgricultura$INDICADOR[dadosAgricultura$INDICADOR == "Area harvested"] <-
      "rea Colhida"
1322 dadosAgricultura$INDICADOR[dadosAgricultura$INDICADOR == "Production"] <-
      "Produto"
1323 dadosAgricultura$INDICADOR[dadosAgricultura$INDICADOR == "Yield"] <-
      "Produtividade"
1324
1325 #traduzindo o nome dos produtos
1326 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Animals live nes"] <- "Animais
      Vivos"
1327 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Asses"] <- "Asno"
1328 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Beehives"] <- "Colmia"
1329 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Buffaloes"] <- "Bfalo"
```

```
1330 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Camelids, other"] <- "Outros
    Cameldeos"
1331 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Camels"] <- "Camelos"
1332 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Cattle"] <- "Gado"
1333 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Chickens"] <- "Frango"
1334 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Ducks"] <- "Pato"
1335 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Geese and guinea fowls"] <-
    "Ganso e Galinha D'Angola"
1336 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Goats"] <- "Cabra"
1337 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Horses"] <- "Cavalo"
1338 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Mules"] <- "Mula"
1339 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Pigeons, other birds"] <-
    "Pombo e Outras Aves"
1340 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Pigs"] <- "Porco"
1341 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Rabbits and hares"] <- "Coelho
    e Lebre"
1342 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Rodents, other"] <- "Outros
    Roedores"
1343 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Sheep"] <- "Ovelha"
1344 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Turkeys"] <- "Peru"
1345 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Beeswax"] <- "Cera de abelha"
1346 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Eggs, hen, in shell"] <- "Ovos
    de Galinha com Casca"
1347 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Eggs, hen, in shell (number)"]
    <- "Ovos de Galinha com Casca"
1348 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Eggs, other bird, in shell"]
    <- "Ovos de Outras Aves com Casca"
1349 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Eggs, other bird, in shell
    (number)"] <- "Ovos de Outras Aves com Casca"
1350 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Hair, horse"] <- "Crina de
    Cavalo"
1351 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Hides, buffalo, fresh"] <-
    "Couro de Bfalo"
1352 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Hides, cattle, fresh"] <-
    "Couro de Gado"
1353 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Honey, natural"] <- "Mel"
1354 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat indigenous, ass"] <-
    "Abate - Animais Selvagens"
1355 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat indigenous, bird nes"] <-
    "Abate - Aves Selvagens"
1356 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat indigenous, buffalo"] <-
    "Abate - Bfalos Selvagens"
```

```
1357 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat indigenous, camel"] <-  
      "Abate - Camelos Selvagens"  
1358 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat indigenous, cattle"] <-  
      "Abate - Bois Selvagens"  
1359 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat indigenous, chicken"] <-  
      "Abate - Frangos Selvagens"  
1360 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat indigenous, duck"] <-  
      "Abate - Patos Selvagens"  
1361 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat indigenous, geese"] <-  
      "Abate - Gansos Selvagens"  
1362 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat indigenous, goat"] <-  
      "Abate - Cabras Selvagens"  
1363 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat indigenous, horse"] <-  
      "Abate - Cavalos Selvagens"  
1364 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat indigenous, mule"] <-  
      "Abate - Mulas Selvagens"  
1365 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat indigenous, other  
      camelids"] <- "Abate - Outros Cameldeos Selvagens"  
1366 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat indigenous, pig"] <-  
      "Abate - Porcos Selvagens"  
1367 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat indigenous, rabbit"] <-  
      "Abate - Coelhoes Selvagens"  
1368 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat indigenous, rodents"] <-  
      "Abate - Roedores Selvagens"  
1369 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat indigenous, sheep"] <-  
      "Abate - Ovelhas Selvagens"  
1370 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat indigenous, turkey"] <-  
      "Abate - Perus Selvagens"  
1371 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat, ass"] <- "Abate -  
      Animais"  
1372 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat, bird nes"] <- "Abate -  
      Aves"  
1373 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat, buffalo"] <- "Abate -  
      Bfalos"  
1374 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat, camel"] <- "Abate -  
      Camelos"  
1375 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat, cattle"] <- "Abate -  
      Bois"  
1376 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat, chicken"] <- "Abate -  
      Frangos"  
1377 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat, duck"] <- "Abate - Patos"
```

```
1378 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat, goat"] <- "Abate -  
    Cabras"  
1379 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat, goose and guinea fowl"]  
    <- "Abate - Gansos e Galinhas D'Angola"  
1380 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat, horse"] <- "Abate -  
    Cavalos"  
1381 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat, mule"] <- "Abate - Mulas"  
1382 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat, other camelids"] <-  
    "Abate - Outros Cameldeos"  
1383 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat, other rodents"] <-  
    "Abate - Roedores"  
1384 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat, pig"] <- "Abate - Porcos"  
1385 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat, rabbit"] <- "Abate -  
    Coelhos"  
1386 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat, sheep"] <- "Abate -  
    Ovelhas"  
1387 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Meat, turkey"] <- "Abate -  
    Perus"  
1388 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Milk, whole fresh buffalo"] <-  
    "Bfala - Leite"  
1389 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Milk, whole fresh camel"] <-  
    "Camelo - Leite"  
1390 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Milk, whole fresh cow"] <-  
    "Vaca - Leite"  
1391 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Milk, whole fresh goat"] <-  
    "Cabra - Leite"  
1392 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Milk, whole fresh sheep  
    Offals, nes"] <- "Ovelha - Leite"  
1393 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Pigeons, other birds"] <-  
    "Pombos e Outras Aves"  
1394 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Silk-worm cocoons, reelable"]  
    <- "Bicho da Seda"  
1395 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Skins, furs"] <- "Couro Animal"  
1396 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Skins, goat, fresh"] <- "Couro  
    de Cabra"  
1397 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Skins, sheep, fresh"] <-  
    "Couro de Ovelhas"  
1398 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Skins, sheep, with wool"] <-  
    "Couro de Ovelha com L"  
1399 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Snails, not sea"] <- "Caracol"  
1400 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Wool, greasy"] <- "Animais - L"  
1401
```

```
1402 #traduzindo os indicados da pecuaria
1403 dadosPecuaria$INDICADOR[dadosPecuaria$INDICADOR == "Production"] <- "Produo"
1404 dadosPecuaria$INDICADOR[dadosPecuaria$INDICADOR == "Milk Animals"] <- "Produo"
1405 dadosPecuaria$INDICADOR[dadosPecuaria$INDICADOR == "Laying"] <- "Produo"
1406 dadosPecuaria$INDICADOR[dadosPecuaria$INDICADOR == "Yield/Carcass Weight"] <-
      "Produtividade"
1407 dadosPecuaria$INDICADOR[dadosPecuaria$INDICADOR == "Prod Popultn"] <- "Produo"
1408 dadosPecuaria$INDICADOR[dadosPecuaria$INDICADOR == "Producing
      Animals/Slaughtered"] <- "Produo"
1409 dadosPecuaria$INDICADOR[dadosPecuaria$INDICADOR == "Yield"] <- "Produtividade"
1410
1411 }, error = function(e) {
1412   print("ERRO NA EXECUCO DA FAO")
1413   writeLines(paste0("Ocorreu um erro na limpeza dos dados da FAO! - ", e),
      paste0(caminhoDownloadArquivos,"erro_limpeza_FAO_", Sys.time(), ".txt"));
1414 })
1415
1416 ##### Limpeza geral dos dados #####
1417
1418 tryCatch({
1419   #removendo registros com valor 0
1420   dadosAgricultura <- dadosAgricultura %>% filter(VALOR > 0)
1421   dadosAgricultura <- dadosAgricultura %>% filter(ANO > 1)
1422   dadosPecuaria <- dadosPecuaria %>% filter(VALOR > 0)
1423   dadosPecuaria <- dadosPecuaria %>% filter(ANO > 1)
1424
1425   #padronizando nome de estados
1426   dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "AC"] <- "Acre"
1427   dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "AM"] <- "Amazonas"
1428   dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "PA"] <- "Par"
1429   dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "TO"] <- "Tocantins"
1430   dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "MA"] <- "Maranhao"
1431   dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "PI"] <- "Piau"
1432   dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "CE"] <- "Cear"
1433   dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "RN"] <- "Rio Grande do
      Norte"
1434   dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "PB"] <- "Paraba"
1435   dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "PE"] <- "Pernambuco"
1436   dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "AL"] <- "Alagoas"
1437   dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "SE"] <- "Sergipe"
1438   dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "BA"] <- "Bahia"
1439   dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "MT"] <- "Mato Grosso"
```

```
1440 dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "MS"] <- "Mato Grosso do
    Sul"
1441 dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "GO"] <- "Goiás"
1442 dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "MG"] <- "Minas Gerais"
1443 dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "ES"] <- "Espírito Santo"
1444 dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "RO"] <- "Rorônia"
1445 dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "RJ"] <- "Rio de Janeiro"
1446 dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "SP"] <- "São Paulo"
1447 dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "DF"] <- "Distrito Federal"
1448 dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "PR"] <- "Paraná"
1449 dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "SC"] <- "Santa Catarina"
1450 dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "RS"] <- "Rio Grande do Sul"
1451 dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "AP"] <- "Amapá"
1452 dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "RR"] <- "Roraima"
1453 dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "BRASIL"] <- "Brasil"
1454 dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "NORTE"] <- "Norte"
1455 dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "NORDESTE"] <- "Nordeste"
1456 dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "SUDESTE"] <- "Sudeste"
1457 dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "CENTRO-OESTE"] <-
    "Centro-Oeste"
1458 dadosAgricultura$ESTADO[dadosAgricultura$ESTADO == "SUL"] <- "Sul"
1459
1460 dadosPecuaria$ESTADO[dadosPecuaria$ESTADO == "BR"] <- "Brasil"
1461
1462 #padronizando nome de produtos
1463 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Abacates"] <- "Abacate"
1464 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Algodão Herbáceo" |
    dadosAgricultura$PRODUTO == "Algodão herbáceo"] <- "Algodão Herbáceo"
1465 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "ALGODO"] <- "Algodão"
1466 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "ALGODO EM CAROÓ"] <-
    "Algodão em Caroo"
1467 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "CAROÓ DE ALGODO"] <-
    "Algodão em Caroo"
1468 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "CAROÓ DE ALGODO BRASIL"]
    <- "Algodão em Caroo"
1469 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "ALGODO EM PLUMA"] <-
    "Algodão em Pluma"
1470 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "AMENDOIM 1 SAFRA" |
    dadosAgricultura$PRODUTO == "Amendoim 1 Safra"] <- "Amendoim (1 Safra)"
1471 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "AMENDOIM 2 SAFRA" |
    dadosAgricultura$PRODUTO == "Amendoim 2 Safra"] <- "Amendoim (2 Safra)"
```



```
1472 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "AMENDOIM TOTAL (1 e 2
      SAFRA)"] <- "Amendoim - Total"
1473 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Amendoim com Casca"] <-
      "Amendoim - Total"
1474 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "ARROZ"] <- "Arroz"
1475 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "ARROZ IRRIGADO"] <-
      "Arroz Irrigado"
1476 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "ARROZ SEQUEIRO"] <-
      "Arroz Sequeiro"
1477 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "AVEIA"] <- "Aveia"
1478 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Batata Inglesa 1 Safra"
      | dadosAgricultura$PRODUTO == "Batata - inglesa (1 Safra)"] <- "Batata
      Inglesa (1 Safra)"
1479 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Batata Inglesa 2 Safra"
      | dadosAgricultura$PRODUTO == "Batata - inglesa (2 Safra)"] <- "Batata
      Inglesa (2 Safra)"
1480 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Batata Inglesa 3 Safra"
      | dadosAgricultura$PRODUTO == "Batata - inglesa (3 Safra)"] <- "Batata
      Inglesa (3 Safra)"
1481 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Caf arbica"] <- "Caf
      Arbica"
1482 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Caf canephora"] <- "Caf
      Canephora"
1483 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Caf total"] <- "Caf
      Total"
1484 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Cana-de-acar" |
      dadosAgricultura$PRODUTO == "CANA-DE-ACAR"] <- "Cana-de-Acar"
1485 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "CANOLA"] <- "Canola"
1486 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Castanha" |
      dadosAgricultura$PRODUTO == "Castanha de Caju"] <- "Castanha-de-Caju"
1487 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Castanha-de-caju"] <-
      "Castanha-de-Caju"
1488 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "CENTEIO"] <- "Centeio"
1489 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "CEVADA"] <- "Cevada"
1490 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Coco-da-baa"] <- "Coco
      da Baa"
1491 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "FEIJO 1 SAFRA" |
      dadosAgricultura$PRODUTO == "Feijo 1 Safra"] <- "Feijo (1 Safra)"
1492 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "FEIJO 2 SAFRA" |
      dadosAgricultura$PRODUTO == "Feijo 2 Safra"] <- "Feijo (2 Safra)"
1493 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "FEIJO 3 SAFRA" |
      dadosAgricultura$PRODUTO == "Feijo 3 Safra"] <- "Feijo (3 Safra)"
```



```
1494 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "FEIJO TOTAL (1, 2 e 3
      SAFRA)"] <- "Feijo - Total"
1495 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "GIRASSOL"] <- "Girassol"
1496 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "MAMONA"] <- "Mamona"
1497 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "MILHO 1 SAFRA" |
      dadosAgricultura$PRODUTO == "Milho 1 Safra"] <- "Milho (1 Safra)"
1498 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "MILHO 2 SAFRA" |
      dadosAgricultura$PRODUTO == "Milho 2 Safra"] <- "Milho (2 Safra)"
1499 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Milho"] <- "Milho -
      Total"
1500 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "MILHO TOTAL (1 e 2
      SAFRA)"] <- "Milho - Total"
1501 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Pimenta-do-reino"] <-
      "Pimenta do Reino"
1502 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "Sisal ou agave"] <-
      "Sisal"
1503 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "SOJA"] <- "Soja"
1504 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "SORGO"] <- "Sorgo"
1505 dadosAgricultura$PRODUTO[dadosAgricultura$PRODUTO == "TRIGO"] <- "Trigo"
1506
1507 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Bois" &
      (dadosPecuaria$INDICADOR == " Animais abatidos" | dadosPecuaria$INDICADOR
      == " Peso total das carcaas")] <- "Abate - Bois"
1508 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Vacas" &
      (dadosPecuaria$INDICADOR == " Animais abatidos" | dadosPecuaria$INDICADOR
      == " Peso total das carcaas")] <- "Abate - Vacas"
1509 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Novilhas" &
      (dadosPecuaria$INDICADOR == " Animais abatidos" | dadosPecuaria$INDICADOR
      == " Peso total das carcaas")] <- "Abate - Novilhas "
1510 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Novilhos" &
      (dadosPecuaria$INDICADOR == " Animais abatidos" | dadosPecuaria$INDICADOR
      == " Peso total das carcaas")] <- "Abate - Novilhos"
1511 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Sunos" &
      (dadosPecuaria$INDICADOR == " Animais abatidos" | dadosPecuaria$INDICADOR
      == " Peso total das carcaas")] <- "Abate - Sunos"
1512 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Frangos" &
      (dadosPecuaria$INDICADOR == " Animais abatidos" | dadosPecuaria$INDICADOR
      == " Peso total das carcaas")] <- "Abate - Frangos"
1513 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Vitelos e vitelas" &
      (dadosPecuaria$INDICADOR == " Animais abatidos" | dadosPecuaria$INDICADOR
      == " Peso total das carcaas")] <- "Abate - Vitelos e Vitelas"
```

```
1514 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Ovos de Galinha" &
      dadosPecuaria$INDICADOR == " Nmero de galinhas poedeiras"] <- "Galinhas
      Poedeiras"
1515 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Animais - L"] <- "L"
1516 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Abate - Frango"] <- "Abate -
      Frangos"
1517 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Abate - Suno"] <- "Abate -
      Sunos"
1518 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Abate - Novilhas "] <- "Abate
      - Novilhas"
1519 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Abate - Vitelos e vitelas"] <-
      "Abate - Vitelos e Vitelas"
1520 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Leite"] <- "Leite de Vaca"
1521 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Bovino"] <- "Boi"
1522 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Mel"] <- "Mel de Abelha"
1523 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Ovos de galinha" |
      dadosPecuaria$PRODUTO == "Ovos"] <- "Ovos de Galinha"
1524 dadosPecuaria$PRODUTO[dadosPecuaria$PRODUTO == "Vaca - Leite"] <- "Vacas
      Leiteiras"
1525
1526 #padronizando nome de indicadores
1527 dadosAgricultura$INDICADOR[dadosAgricultura$INDICADOR == "rea Colhida" |
      dadosAgricultura$INDICADOR == " rea colhida"] = "rea Colhida"
1528 dadosAgricultura$INDICADOR[dadosAgricultura$INDICADOR == " rea plantada"] <-
      "rea Plantada"
1529 dadosAgricultura$INDICADOR[dadosAgricultura$INDICADOR == " Produo"] <- "Produo"
1530 dadosAgricultura$INDICADOR[dadosAgricultura$INDICADOR == " Rendimento mdio"]
      <- "Produtividade"
1531
1532 dadosPecuaria$INDICADOR[dadosPecuaria$INDICADOR == " Animais abatidos"] <-
      "Rebanho"
1533 dadosPecuaria$INDICADOR[dadosPecuaria$INDICADOR == " Nmero de galinhas
      poedeiras"] <- "Rebanho"
1534 dadosPecuaria$INDICADOR[dadosPecuaria$INDICADOR == " Peso total das carcaas"]
      <- "Produo"
1535 dadosPecuaria$INDICADOR[dadosPecuaria$INDICADOR == " Quantidade de leite cru,
      resfriado ou no, adquirido"] <- "Produo"
1536 dadosPecuaria$INDICADOR[dadosPecuaria$INDICADOR == " Quantidade de ovos
      produzidos"] <- "Produo"
1537 dadosPecuaria$INDICADOR[dadosPecuaria$INDICADOR == "Stocks"] <- "Rebanho"
1538 dadosPecuaria$INDICADOR[dadosPecuaria$INDICADOR == "Produo" &
      dadosPecuaria$UNIDADE == "Mil Cabeas"] <- "Rebanho"
```

```
1539
1540 #removendo registros das de indicadores diferentes dos pre definidos
1541 dadosAgricultura <- dadosAgricultura %>% filter(INDICADOR == "rea Colhida" |
      INDICADOR == "rea Plantada" | INDICADOR == "Produo" | INDICADOR ==
      "Produtividade")
1542
1543 #padronizando nome de unidades
1544 dadosAgricultura$UNIDADE[dadosAgricultura$UNIDADE == "kg/ha"] <- "Kg/Hectare"
1545 dadosAgricultura$UNIDADE[dadosAgricultura$UNIDADE == "mil hectares"] <- "Mil
      Hectares"
1546 dadosAgricultura$UNIDADE[dadosAgricultura$UNIDADE == "mil toneladas"] <- "Mil
      Toneladas"
1547
1548 dadosPecuaria$UNIDADE[dadosPecuaria$UNIDADE == "Mil dzias"] <- "Mil Dzias"
1549 dadosPecuaria$UNIDADE[dadosPecuaria$UNIDADE == "Mil litros"] <- "Mil Litros"
1550 dadosPecuaria$UNIDADE[dadosPecuaria$UNIDADE == "Mil unidades"] <- "Mil
      Unidades"
1551
1552 }, error = function(e) {
1553   print("ERRO NA EXECUCO DA LIMPEZA GERAL")
1554   writeLines(paste0("Ocorreu um erro na limpeza geral dos dados! - ", e),
      paste0(caminhoDownloadArquivos,"erro limpeza geral ", Sys.time(),
      ".txt"));
1555 })
1556
1557 #salvando o xlsx
1558 write_xlsx(dadosPecuaria, paste0(caminhoDownloadArquivos,
      "dadosPecuaria.xlsx"));
1559 write_xlsx(dadosAgricultura, paste0(caminhoDownloadArquivos,
      "dadosAgricultura.xlsx"));
1560
1561 deployApp(appDir = caminhoDownloadArquivos, appFiles = c("app.R",
      "dadosAgricultura.xlsx", "dadosPecuaria.xlsx"), appName =
      "indicadoresAgro")
```

APÊNDICE E – *SCRIPT* DA FERRAMENTA EM R DE ANÁLISE DOS DADOS

```

1 library(shiny)
2 library(ggplot2)
3 require(tidyverse)
4 library(readxl)
5 library(plotly)
6 library(shinydashboard)
7 library(leaflet)
8 dadosAgricultura <- read_excel("dadosAgricultura.xlsx")
9 dadosPecuaria <- read_excel("dadosPecuaria.xlsx")
10 dadosAgricultura <- dadosAgricultura %>%
11   mutate(ANO = as.numeric(substr(ANO, 1,4)),
12          VALOR = as.numeric(VALOR))
13 dadosPecuaria <- dadosPecuaria %>%
14   mutate(ANO = as.numeric(substr(ANO, 1,4)),
15          VALOR = as.numeric(VALOR))
16
17 abaAgricultura <- function() {
18   #montando barra lateral com filtros para o usuario
19   sidebarLayout(
20     sidebarPanel(
21       helpText(" INSTRUÇÕES : "),
22       helpText("Selecione ao menos uma fonte, um produto e um estado ou regio
23         para que os dados sejam gerados no gráfico"),
24       sliderInput("range",
25         "Anos",
26         min = min(dadosAgricultura$ANO, na.rm = T),
27         max = max(dadosAgricultura$ANO, na.rm = T),
28         value = c(min(dadosAgricultura$ANO, na.rm = T),
29           max(dadosAgricultura$ANO, na.rm = T))),
30       checkboxGroupInput("Fonte", "Fontes:",
31         choices = c(unique(dadosAgricultura$FONTE)),
32         selected = ""),
33       selectInput("Produtos", "Produtos",
34         c(sort(unique(dadosAgricultura$PRODUTO))), multiple = TRUE),
35       selectInput("Estados", "Estados",
36         c(sort(unique(dadosAgricultura$ESTADO))), multiple = TRUE)

```

```
33     ),
34
35     # montando painel principal com os graficos e estatsticas
36     mainPanel(
37         wellPanel(
38             align="center",
39             h4(textOutput("tituloProducao")),
40             splitLayout(
41                 style = "border: 1px solid black;",
42                 plotlyOutput("grafProducao")
43             ),
44             splitLayout(
45                 style = "border: 1px solid black;",
46                 plotlyOutput("boxplotProducao"),
47                 tableOutput("resumoProducao")
48             )
49         ),
50         wellPanel(
51             align="center",
52             h4(textOutput("tituloProdutividade")),
53             splitLayout(
54                 style = "border: 1px solid black;",
55                 plotlyOutput("grafProdutividade")
56             ),
57             splitLayout(
58                 style = "border: 1px solid black;",
59                 plotlyOutput("boxplotProdutividade"),
60                 tableOutput("resumoProdutividade")
61             )
62         ),
63         wellPanel(
64             align="center",
65             h4(textOutput("tituloAreaPlantada")),
66             splitLayout(
67                 style = "border: 1px solid black;",
68                 plotlyOutput("grafAreaPlantada")
69             ),
70             splitLayout(
71                 style = "border: 1px solid black;",
72                 plotlyOutput("boxplotAreaPlantada"),
73                 tableOutput("resumoAreaPlantada")
74             )
75         )
76     )
77 }
```

```

75     ),
76     wellPanel(
77       align="center",
78       h4(textOutput("tituloAreaColhida")),
79       splitLayout(
80         style = "border: 1px solid black;",
81         plotlyOutput("grafAreaColhida")
82       ),
83       splitLayout(
84         style = "border: 1px solid black;",
85         plotlyOutput("boxplotAreaColhida"),
86         tableOutput("resumoAreaColhida")
87       )
88     )
89   )
90 )
91 }
92
93 abaPecuaria <- function() {
94   #montando barra lateral com filtros para o usuario
95   sidebarLayout(
96     sidebarPanel(
97       helpText(" INSTRUES : "),
98       helpText("Selecione ao menos uma fonte, um produto e um estado ou regio
99         para que os dados sejam gerados no grfico"),
100       sliderInput("rangePecuaria",
101         "Anos",
102         min = min(dadosPecuaria$ANO, na.rm = T),
103         max = max(dadosPecuaria$ANO, na.rm = T),
104         value = c(min(dadosPecuaria$ANO, na.rm = T),
105           max(dadosPecuaria$ANO, na.rm = T))),
106       checkboxGroupInput("FontePecuaria", "Fontes:",
107         choices = c(sort(unique(dadosPecuaria$FONTE))),
108         selected = ""),
109       selectInput("ProdutosPecuaria", "Produtos",
110         c(sort(unique(dadosPecuaria$PRODUTO))), multiple = TRUE),
111       selectInput("EstadosPecuaria", "Estados",
112         c(sort(unique(dadosPecuaria$ESTADO))), multiple = TRUE)
113     ),
114     # montando painel principal com os graficos e estatisticas
115     mainPanel(

```

```
115     wellPanel(  
116         align="center",  
117         h4(textOutput("tituloRebanhoPecuaria")),  
118         splitLayout(  
119             style = "border: 1px solid black;",  
120             plotlyOutput("grafRebanhoPecuaria")  
121         ),  
122         splitLayout(  
123             style = "border: 1px solid black;",  
124             plotlyOutput("boxplotRebanhoPecuaria"),  
125             tableOutput("resumoRebanhoPecuaria")  
126         )  
127     ),  
128     wellPanel(  
129         align="center",  
130         h4(textOutput("tituloProducaoPecuaria")),  
131         splitLayout(  
132             style = "border: 1px solid black;",  
133             plotlyOutput("grafProducaoPecuaria")  
134         ),  
135         splitLayout(  
136             style = "border: 1px solid black;",  
137             plotlyOutput("boxplotProducaoPecuaria"),  
138             tableOutput("resumoProducaoPecuaria")  
139         )  
140     ),  
141     wellPanel(  
142         align="center",  
143         h4(textOutput("tituloProdutividadePecuaria")),  
144         splitLayout(  
145             style = "border: 1px solid black;",  
146             plotlyOutput("grafProdutividadePecuaria")  
147         ),  
148         splitLayout(  
149             style = "border: 1px solid black;",  
150             plotlyOutput("boxplotProdutividadePecuaria"),  
151             tableOutput("resumoProdutividadePecuaria")  
152         )  
153     )  
154 )  
155 )  
156 }
```



```

157
158 abaMetadados <- function() {
159   fluidRow(
160     tags$div(class = "header", style="width: 98%;border: 1px solid #dddddd;
161       background-color: #f5f5f5; margin: 1%;", checked = NA,
162       tags$style(HTML("table {
163         font-family: arial, sans-serif;
164         border-collapse: collapse;
165         width: 100%;
166       }
167       td, th {
168         border: 1px solid black;
169         text-align: left;
170         padding: 8px;
171       }
172       tr:nth-child(even) {
173         background-color: #dddddd;
174       }"))),
175     tags$p(HTML("<table style='width: 80%; margin: 10%'>
176       <tr>
177       <th style='width: 30%; text-align: center'>Elemento</th>
178       <th style='width: 70%'>Valor</th>
179     </tr>
180     <tr>
181       <td style='text-align: center'>Titulo</td>
182       <td>Indicadores de Produo do Agronegcio
183       Brasileiro</td>
184     </tr>
185     <tr>
186       <td style='text-align: center'>Assunto</td>
187       <td>Agronegcio, Indicadores de Produo, Visualizao da
188       Informao </td>
189     </tr>
190     <tr>
191       <td style='text-align: center'>Relaes</td> <td>CONAB -
192       Portal de Informaes Agropecurias Observatrio
193       Agrcola <br>EMBRAPA - AGROPENSA </td>
194     </tr>
195   )
196 }
```

```

191         <td style='text-align: center'>Fonte</td> <td> CONAB,
192             EMBRAPA, FAO, IBGE</td>
193     </tr>
194     <tr>
195         <td style='text-align:
196             center'>Cobertura</td><td>1000047</td>
197     </tr>
198     <tr>
199         <td style='text-align: center'>Tipo</td> <td>Dados
200             relacionados a produo do agronegcio brasileiro</td>
201     </tr>
202     <tr>
203         <td style='text-align: center'>Descrigo</td> <td>A
204             ferramenta tem como objetivo disponibilizar dados do
205             agronegcio brasileiro atravs de grficos e filtros
206             interativos</td>
207     </tr>
208     <tr>
209         <td style='text-align: center'>Criador</td> <td>Jhonny
210             Ivair de Lima Maciel</td>
211     </tr>
212     <tr>
213         <td style='text-align: center'>Colaborador</td>
214             <td>Prof. Dr. Egon Walter Wildauer</td>
215     </tr>
216     <tr>
217         <td style='text-align: center'>Data</td> <td>".
218             format(Sys.time()),"</td>
219     </tr>
220     <tr>
221         <td style='text-align: center'>Formato</td>
222             <td>Ferramenta web em R utilizando a biblioteca
223             Shiny</td>
224     </tr>
225     <tr>
226         <td style='text-align: center'>Idioma</td> <td>pt</td>
227     </tr>
228 </table>"))
229 )
230 }
231

```

```

222 # Define UI for application that draws a histogram
223 ui <- fluidPage(
224   titlePanel("Indicadores de Produo do Agronegcio Brasileiro"),
225   tabsetPanel(
226     tabPanel("Agricultura", abaAgricultura()),
227     tabPanel("Pecuaria", abaPecuaria()),
228     tabPanel("Metadados", abaMetadados())
229   )
230 )
231
232 # Define server logic required to draw a histogram
233 server <- function(input, output, session) {
234   #graficos da parte de agricultura
235   dados <- reactive({
236     dados = dadosAgricultura %>%
237       mutate(Dcada = factor(sprintf("%02d", 10 * (as.integer(ANO/10) %% 10)),
238         levels = c("60", "70", "80", "90", "00", "10", "20"))) %>%
239       filter(between(ANO, min(input$range), max(input$range))
240         & (ESTADO %in% input$Estados)
241         & (PRODUTO %in% input$Produtos)
242         & (FONTE %in% input$Fonte))
243   })
244
245   #titulo PRODUCAO
246   output$tituloProducao <- renderText({
247     "Dados de Produo "
248   })
249
250   #grafico PRODUCAO
251   output$grafProducao <- renderPlotly({
252     tryCatch({
253       grafProducao = dados() %>%
254         filter(INDICADOR == ' Produo ') %>%
255         ggplot(aes(x = ANO, y = VALOR, color=ESTADO, linetype = PRODUTO,
256           shape=FONTE)) +
257         ggtitle(' Produo - Mil Toneladas') +
258         geom_line() +
259         geom_point() +
260         theme_bw()
261       ggplotly(grafProducao) %>% config(displayModeBar = F)
262     },
263     error=function(e) {

```

```
263     ggplot()+
264     ggtitle('Nenhum dado para reproduzir no gráfico, altere os filtros!')
265   }
266 )
267 })
268
269 #estatísticas PRODUÇÃO
270 output$resumoProducao <- renderTable({
271   tryCatch({
272     dados() %>%
273       filter(INDICADOR == 'Produto') %>%
274       group_by(Dcada) %>%
275       summarise(Media = mean(VALOR, na.rm = T),
276                 "Desvio Padrao" = sd(VALOR, na.rm = T),
277                 Minimo = min(VALOR, na.rm = T),
278                 Maximo = max(VALOR, na.rm = T))
279   }, error=function(e){})
280 })
281
282 #boxplot PRODUÇÃO
283 output$boxplotProducao <- renderPlotly({
284   tryCatch({
285     dados() %>%
286       filter(INDICADOR == 'Produto') %>%
287       ggplot(aes(x = Dcada, y = VALOR)) +
288       ggtitle('Boxplot - Produto') +
289       geom_boxplot() +
290       geom_jitter(aes(shape=FONTE, color=ESTADO, text = PRODUTO)) +
291       theme(legend.position = "none")
292   },error=function(e) {
293     ggplot()+
294     ggtitle('Nenhum dado para reproduzir nos boxplots, altere os filtros!')
295   })
296 })
297
298 #título PRODUTIVIDADE
299 output$tituloProdutividade <- renderText({
300   "Dados de Produtividade"
301 })
302
303 #gráfico PRODUTIVIDADE
304 output$grafProdutividade <- renderPlotly({
```

```
305   tryCatch(  
306     {  
307       graficoProdutividade = dados() %>%  
308         filter(INDICADOR == 'Produtividade') %>%  
309         ggplot() +  
310         ggtitle('Produtividade - Kg/Hectare') +  
311         geom_line(aes(x = ANO, y = VALOR, color=ESTADO, linetype = PRODUTO,  
312           shape=FONTE)) +  
313         geom_point(aes(x = ANO, y = VALOR, shape=FONTE, color=ESTADO,linetype  
314           = PRODUTO)) +  
315         theme_bw()  
316         ggplotly(graficoProdutividade) %>% config(displayModeBar = F)  
317     },  
318     error=function(e) {  
319       ggplot()+  
320       ggtitle('Nenhum dado para reproduzir no grfico, altere os filtros!')  
321     }  
322   )  
323 })  
324  
325 #estatisticas PRODUTIVIDADE  
326 output$resumoProdutividade <- renderTable({  
327   tryCatch({  
328     dados() %>%  
329     filter(INDICADOR == 'Produtividade') %>%  
330     group_by(Dcada) %>%  
331     summarise(Media = mean(VALOR, na.rm = T),  
332       "Desvio Padrao" = sd(VALOR, na.rm = T),  
333       Minimo = min(VALOR, na.rm = T),  
334       Maximo = max(VALOR, na.rm = T))  
335   },error=function(e) {})  
336 })  
337  
338 #boxplot PRODUTIVIDADE  
339 output$boxplotProdutividade<- renderPlotly({  
340   tryCatch({  
341     dados() %>%  
342     filter(INDICADOR == 'Produtividade') %>%  
343     ggplot(aes(x = Dcada, y = VALOR)) +  
344     ggtitle('Boxplot - Produtividade') +  
345     geom_boxplot() +  
346     geom_jitter(aes(shape=FONTE, color=ESTADO, text = PRODUTO)) +
```

```
345     theme(legend.position = "none")
346   },error=function(e) {
347     ggplot()+
348       ggtitle('Nenhum dado para reproduzir nos boxplots, altere os filtros!')
349   })
350 })
351
352 #titulo AREA PLANTADA
353 output$tituloAreaPlantada <- renderText({
354   "Dados de rea Plantada"
355 })
356
357 #grfico AREA PLANTADA
358 output$grafAreaPlantada <- renderPlotly({
359   tryCatch({
360     graficoAreaPlantada = dados() %>%
361       filter(INDICADOR == 'rea Plantada') %>%
362       ggplot(aes(x = ANO, y = VALOR, color=ESTADO, linetype = PRODUTO,
363         shape=FONTE)) +
364       ggtitle('rea Plantada - Mil Hectares') +
365       geom_line() +
366       geom_point() +
367       theme_bw()
368     ggplotly(graficoAreaPlantada) %>% config(displayModeBar = F)
369   },
370   error=function(e) {
371     ggplot()+
372       ggtitle('Nenhum dado para reproduzir no grafico, altere os filtros!')
373   })
374 })
375
376 #estatsticas AREA PLANTADA
377 output$resumoAreaPlantada <- renderTable({
378   tryCatch({
379     dados() %>%
380       filter(INDICADOR == 'rea Plantada') %>%
381       group_by(Dcada) %>%
382       summarise(Media = mean(VALOR, na.rm = T),
383         "Desvio Padrao" = sd(VALOR, na.rm = T),
384         Minimo = min(VALOR, na.rm = T),
385         Maximo = max(VALOR, na.rm = T))
386   },error=function(e) {})
```

```
386   })
387
388   #boxplot AREA PLANTADA
389   output$boxplotAreaPlantada<- renderPlotly({
390     tryCatch({
391       dados() %>%
392         filter(INDICADOR == 'rea Plantada') %>%
393         ggplot(aes(x = Dcada, y = VALOR)) +
394         ggtitle('Boxplot - rea Plantada') +
395         geom_boxplot() +
396         geom_jitter(aes(shape=FONTE, color=ESTADO, text = PRODUTO)) +
397         theme(legend.position = "none")
398     },error=function(e) {
399       ggplot()+
400       ggtitle('Nenhum dado para reproduzir nos boxplots, altere os filtros!')
401     })
402   })
403
404   #titulo AREA COLHIDA
405   output$tituloAreaColhida <- renderText({
406     "Dados de rea Colhida"
407   })
408
409   # grfico AREA COLHIDA
410   output$grafAreaColhida <- renderPlotly({
411     tryCatch({
412       graficoAreaColhida = dados() %>%
413         filter(INDICADOR == 'rea Colhida') %>%
414         ggplot(aes(x = ANO, y = VALOR, color=ESTADO, linetype = PRODUTO,
415           shape=FONTE)) +
416         ggtitle('rea Colhida - Mil Hectares') +
417         geom_line() +
418         geom_point() +
419         theme_bw()
420       ggplotly(graficoAreaColhida) %>% config(displayModeBar = F)
421     },error=function(e) {
422       ggplot()+
423       ggtitle('Nenhum dado para reproduzir no grfico, altere os filtros!')
424     })
425   })
426
427   #estatsticas AREA COLHIDA
```



```

427   output$resumoAreaColhida <- renderTable({
428     tryCatch({
429       dados() %>%
430         filter(INDICADOR == 'rea Colhida') %>%
431         group_by(Dcada) %>%
432         summarise(Media = mean(VALOR, na.rm = T),
433                   "Desvio Padrao" = sd(VALOR, na.rm = T),
434                   Minimo = min(VALOR, na.rm = T),
435                   Maximo = max(VALOR, na.rm = T))
436     },error=function(e) {})
437   })
438
439   #boxplot AREA COLHIDA
440   output$boxplotAreaColhida<- renderPlotly({
441     tryCatch({
442       dados() %>%
443         filter(INDICADOR == 'rea Colhida') %>%
444         ggplot(aes(x = Dcada, y = VALOR)) +
445         ggtitle('Boxplot - rea Colhida') +
446         geom_boxplot() +
447         geom_jitter(aes(shape=FONTE, color=ESTADO, text = PRODUTO)) +
448         theme(legend.position = "none")
449     },error=function(e) {
450       ggplot()+
451       ggtitle('Nenhum dado para reproduzir nos boxplots, altere os filtros!')
452     })
453   })
454
455   #graficos da parte de pecuaria
456   dadosP <- reactive({
457     dadosP = dadosPecuaria %>%
458       mutate(Dcada = factor(sprintf("%02d", 10 * (as.integer(ANO/10) %% 10)),
459                                levels = c("60","70","80","90","00","10","20"))) %>%
460       filter(between(ANO, min(input$rangePecuaria), max(input$rangePecuaria))
461             & (ESTADO %in% input$EstadosPecuaria)
462             & (PRODUTO %in% input$ProdutosPecuaria)
463             & (FONTE %in% input$FontePecuaria))
464   })
465
466   #titulo REBANHO Pecuaria
467   output$tituloRebanhoPecuaria <- renderText({
468     "Dados de Rebanho"

```

```
469   })
470
471   #grafico REBANHO Pecuaria
472   output$grafRebanhoPecuaria <- renderPlotly({
473     tryCatch({
474       grafRebanhoPecuaria = dadosP() %>%
475         filter(INDICADOR == 'Rebanho') %>%
476         ggplot(aes(x = ANO, y = VALOR, color=ESTADO, linetype = PRODUTO,
477           shape=FONTE)) +
478         ggtitle('Rebanho - Mil Cabeas') +
479         geom_line() +
480         geom_point() +
481         theme_bw()
482       ggplotly(grafRebanhoPecuaria) %>% config(displayModeBar = F)
483     },
484     error=function(e) {
485       ggplot()+
486       ggtitle('Nenhum dado para reproduzir no grfico, altere os filtros!')
487     }
488   })
489
490   #estatsticas REBANHO Pecuaria
491   output$resumoRebanhoPecuaria <- renderTable({
492     tryCatch({
493       dadosP() %>%
494       filter(INDICADOR == 'Rebanho') %>%
495       group_by(Dcada) %>%
496       summarise(Media = mean(VALOR, na.rm = T),
497         "Desvio Padrao" = sd(VALOR, na.rm = T),
498         Minimo = min(VALOR, na.rm = T),
499         Maximo = max(VALOR, na.rm = T))
500     }, error=function(e){})
501   })
502
503   #boxplot REBANHO Pecuaria
504   output$boxplotRebanhoPecuaria <- renderPlotly({
505     tryCatch({
506       dadosP() %>%
507       filter(INDICADOR == 'Rebanho') %>%
508       ggplot(aes(x = Dcada, y = VALOR)) +
509       ggtitle('Boxplot - Rebanho') +
```

```
510     geom_boxplot() +
511     geom_jitter(aes(shape=FONTE, color=ESTADO, text = PRODUTO)) +
512     theme(legend.position = "none")
513   },error=function(e) {
514     ggplot()+
515     ggtitle('Nenhum dado para reproduzir nos boxplots, altere os filtros!')
516   })
517 })
518
519 #titulo PRODUCAO Pecuaria
520 output$tituloProducaoPecuaria <- renderText({
521   "Dados de Produto "
522 })
523
524 #grafico PRODUCAO Pecuaria
525 output$grafProducaoPecuaria <- renderPlotly({
526   tryCatch({
527     grafProducaoPecuaria = dadosP() %>%
528       filter(INDICADOR == 'Produto') %>%
529       ggplot(aes(x = ANO, y = VALOR, color=ESTADO, linetype = PRODUTO,
530                 shape=FONTE)) +
531       ggtitle('Produto - Mil Toneladas/Mil Litros/Mil Dzas/Mil Unidades') +
532       geom_line() +
533       geom_point() +
534       theme_bw()
535     ggplotly(grafProducaoPecuaria) %>% config(displayModeBar = F)
536   },
537   error=function(e) {
538     ggplot()+
539     ggtitle('Nenhum dado para reproduzir no grafico, altere os filtros!')
540   })
541 })
542
543 #estatisticas PRODUCAO Pecuaria
544 output$resumoProducaoPecuaria <- renderTable({
545   tryCatch({
546     dadosP() %>%
547       filter(INDICADOR == 'Produto') %>%
548       group_by(Dcada) %>%
549       summarise(Media = mean(VALOR, na.rm = T),
550                 "Desvio Padrao" = sd(VALOR, na.rm = T),
```

```
551         Minimo = min(VLOR, na.rm = T),
552         Maximo = max(VLOR, na.rm = T))
553     }, error=function(e){})
554 })
555
556 #boxplot PRODUCAO Pecuaria
557 output$boxplotProducaoPecuaria <- renderPlotly({
558     tryCatch({
559         dadosP() %>%
560             filter(INDICADOR == 'Produto') %>%
561             ggplot(aes(x = Dcada, y = VLOR)) +
562             ggtitle('Boxplot - Produto') +
563             geom_boxplot() +
564             geom_jitter(aes(shape=FONTE, color=ESTADO, text = PRODUTO)) +
565             theme(legend.position = "none")
566     },error=function(e) {
567         ggplot()+
568         ggtitle('Nenhum dado para reproduzir nos boxplots, altere os filtros!')
569     })
570 })
571
572 #titulo PRODUTIVIDADE Pecuaria
573 output$tituloProdutividadePecuaria <- renderText({
574     "Dados de Produtividade"
575 })
576
577 #grafico PRODUTIVIDADE Pecuaria
578 output$grafProdutividadePecuaria <- renderPlotly({
579     tryCatch(
580         {
581             graficoProdutividadePecuaria = dadosP() %>%
582                 filter(INDICADOR == 'Produtividade') %>%
583                 ggplot() +
584                 ggtitle('Produtividade - Kg/Animal') +
585                 geom_line(aes(x = ANO, y = VLOR, color=ESTADO, linetype = PRODUTO,
586                             shape=FONTE)) +
587                 geom_point(aes(x = ANO, y = VLOR, shape=FONTE, color=ESTADO,linetype
588                             = PRODUTO)) +
589                 theme_bw()
590             ggplotly(graficoProdutividadePecuaria) %>% config(displayModeBar = F)
591         },
592         error=function(e) {
```

```
591     ggplot()+
592       ggtitle('Nenhum dado para reproduzir no gráfico, altere os filtros!')
593   }
594 )
595 })
596
597 #estatísticas PRODUTIVIDADE Pecuaria
598 output$resumoProdutividadePecuaria <- renderTable({
599   tryCatch({
600     dadosP() %>%
601       filter(INDICADOR == 'Produtividade') %>%
602       group_by(Dcada) %>%
603       summarise(Media = mean(VALOR, na.rm = T),
604                 "Desvio Padrao" = sd(VALOR, na.rm = T),
605                 Minimo = min(VALOR, na.rm = T),
606                 Maximo = max(VALOR, na.rm = T))
607   },error=function(e) {})
608 })
609
610 #boxplot PRODUTIVIDADE Pecuaria
611 output$boxplotProdutividadePecuaria <- renderPlotly({
612   tryCatch({
613     dadosP() %>%
614       filter(INDICADOR == 'Produtividade') %>%
615       ggplot(aes(x = Dcada, y = VALOR)) +
616       ggtitle('Boxplot - Produtividade') +
617       geom_boxplot() +
618       geom_jitter(aes(shape=FONTE, color=ESTADO, text = PRODUTO)) +
619       theme(legend.position = "none")
620   },error=function(e) {
621     ggplot()+
622       ggtitle('Nenhum dado para reproduzir nos boxplots, altere os filtros!')
623   })
624 })
625 }
626
627 # rodar ferramenta
628 shinyApp(ui = ui, server = server)
```

APÊNDICE F – COMPARAÇÕES DOS RESULTADOS DE APRESENTADOS NO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - SOJA

Tabela 11 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 e DA FERRAMENTA - ÁREA COLHIDA (MIL HECTARES) DE SOJA EM MATO GROSSO

Ano	Sousa, Bonjour e Figueiredo (2006)	IBGE (%)	EMBRAPA (%)
1990	1.527,75	-	1.527,75 (0%)
1991	1.164,58	-	1.164,58 (0%)
1992	1.453,70	-	1.453,70 (0%)
1993	1.678,53	-	1.678,53 (0%)
1994	2.022,95	-	2.022,95 (0%)
1995	2.322,82	-	2.322,82 (0%)
1996	1.956,14	-	1.956,14 (0%)
1997	2.192,51	-	2.192,51 (0%)
1998	2.643,38	-	2.643,38 (0%)
1999	2.635,01	-	2.635,01 (0%)
2000	2.906,44	-	2.906,44 (0%)
2001	3.121,35	-	3.121,35 (0%)
2002	3.818,23	-	3.818,23 (0%)
2003	4.413,27	-	4.413,27 (0%)
2004	5.263,42	-	5.263,42 (0%)
2005	6.106,65	-	6.106,65 (0%)
2006	5.808,67	5808,67 (0%)	5.811,90 (0,05%)

Tabela 12 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - PRODUÇÃO (MIL TONELADAS) DE SOJA EM MATO GROSSO

Ano	Sousa, Bonjour e Figueiredo (2006)	IBGE (%)	EMBRAPA (%)	CONAB (%)
1990	3.064,715	-	3.064,715 (0%)	2.900,8 (-5,35%)
1991	2.738,41	-	2.738,41 (0%)	2.607 (-4,80%)
1992	3.642,743	-	3.642,743 (0%)	3.484,8 (-4,34%)
1993	4.118,726	-	4.118,726 (0%)	4.197,8 (1,92%)
1994	5.319,793	-	5.319,793 (0%)	4.970 (-6,58%)
1995	5.491,426	-	5.491,426 (0%)	5.440,1 (-0,93%)
1996	5.032,921	-	5.032,921 (0%)	4.686,8 (-6,88%)
1997	6.060,882	-	6.060,882 (0%)	5.721,3 (-5,60%)
1998	7.228,085	-	7.228,052 (0%)	7.150 (-1,08%)
1999	7.473,028	-	7.473,028 (0%)	7.134,4 (-4,53%)
2000	8.774,47	-	8.774,47 (0%)	8.801,2 (0,30%)
2001	9.533,286	-	9.533,286 (0%)	9.640,8 (1,13%)
2002	11.684,885	-	11.684,885 (0%)	11.733 (0,41%)
2003	12.965,983	-	12.965,983 (0%)	12.949,4 (-0,13%)
2004	14.517,912	-	14.517,912 (0%)	15.008,8 (3,38%)
2005	17.761,444	-	17.761,444 (0%)	17.937,1 (0,99%)
2006	15.586,887	15586,89 (0%)	15.594,221 (0%)	16.700,4 (-7,14%)

Tabela 13 – COMPARAÇÃO DE RESULTADOS E DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - PRODUTIVIDADE (KG/HA) DE SOJA EM MATO GROSSO

Ano	Sousa, Bonjour e Figueiredo (2006)	IBGE (%)	EMBRAPA (%)	CONAB (%)
1990	2.006	-	2.006 (0%)	1.930 (-3,79%)
1991	2.351	-	2.351 (0%)	2.370 (0,81%)
1992	2.505	-	2.505 (0%)	2.400 (-4,19%)
1993	2.454	-	2.453 (-0,05%)	2.450 (0,16%)
1994	2.630	-	2.629 (-0,04%)	2.490 (-5,32%)
1995	2.364	-	2.364 (0%)	2.370 (0,25%)
1996	2.573	-	2.572 (-0,04%)	2.460 (-4,39%)
1997	2.764	-	2.764 (0%)	2.730 (-1,23%)
1998	2.734	-	2.734 (0%)	2.750 (0,59%)
1999	2.836	-	2.836 (0%)	2.800 (-1,27%)
2000	3.018	-	3.018 (0%)	3.030 (0,40%)
2001	3.054	-	3.054 (0%)	3.090 (1,18%)
2002	3.060	-	3.060 (0%)	3.045 (-0,49%)
2003	2.937	-	2.937 (0%)	2.930 (-0,24%)
2004	2.758	-	2.758 (0%)	2.864 (3,84%)
2005	2.909	-	2.908 (-0,04%)	2.938 (1%)
2006	2.683	2.683 (0%)	2.683 (0%)	2.695 (0,45%)

Tabela 14 – COMPARAÇÃO DE RESULTADOS E DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - ÁREA COLHIDA (MIL HECTARES) DE SOJA NO BRASIL

Ano	Sousa, Bonjour e Figueiredo (2006)	IBGE (%)	EMBRAPA (%)	FAO (%)
1990	11.551,40	-	11.487,30 (-0,6%)	11.487,30 (-0,6%)
1991	9.742,50	-	9.616,65 (-1,3%)	9.616,65 (-1,3%)
1992	9.582,20	-	9.441,39 (-1,5%)	9.441,39 (-1,5%)
1993	10.717,00	-	10.635,33 (-0,8%)	10.635,33 (-0,8%)
1994	11.501,70	-	11.525,41 (0,2%)	11.525,41 (0,2%)
1995	11.678,70	-	11.675,01 (0%)	11.675,01 (0%)
1996	10.299,47	-	10.299,47 (0%)	10.299,47 (0%)
1997	11.486,48	-	11.486,48 (0%)	11.486,48 (0%)
1998	13.303,66	-	13.303,66 (0%)	13.303,66 (0%)
1999	13.061,41	-	13.061,41 (0%)	13.061,41 (0%)
2000	13.656,77	-	13.656,77 (0%)	13.656,77 (0%)
2001	13.985,10	-	13.985,10 (0%)	13.985,10 (0%)
2002	16.359,44	-	16.359,44 (0%)	16.359,44 (0%)
2003	18.254,77	-	18.524,77 (1,5%)	18.524,77 (1,5%)
2004	21.538,99	-	21.538,99 (0,%)	21.538,99 (0%)
2005	22.917,01	-	22.948,87 (0,1%)	22.948,87 (0,1%)
2006	22.006,68	21.958,080 (-0,2%)	22.047,35 (0,2%)	22.047,35 (0,2%)

Tabela 15 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - PRODUÇÃO (MIL TONELADAS) DE SOJA NO BRASIL

Ano	Sousa, Bonjour e Figueiredo (2006)	IBGE (%)	EMBRAPA (%)	CONAB (%)	FAO (%)
1990	20.101,30	-	19.897,8 (-1,0%)	20.101,3 (0%)	19.897,80 (-1,0%)
1991	15.394,50	-	14.937,81 (-3,0%)	15.394,5 (0%)	14.937,81 (-3,0%)
1992	19.418,60	-	19.214,71 (-1,0%)	19.418,6 (0%)	19.214,70 (-1,1%)
1993	23.042,10	-	22.590,98 (-2,0%)	23.042,1 (0%)	22.590,98 (-2,0%)
1994	25.059,20	-	24.931,83 (-0,5%)	25.059,2 (0%)	24.931,83 (-0,5%)
1995	25.934,10	-	25.682,64 (-1,0%)	25.934,1 (0%)	25.682,64 (-1,0%)
1996	23.166,87	-	23.166,87 (0%)	23.189,7 (0,1%)	23.166,87 (0%)
1997	26.392,64	-	26.392,64 (0%)	26.160,0 (-0,88%)	26.392,64 (0%)
1998	31.307,44	-	31.307,44 (0%)	31.369,9 (0,2%)	31.307,44 (0%)
1999	30.978,48	-	30.987,48 (0%)	30.765,0 (-0,69%)	30.987,48 (0%)
2000	38.820,83	-	32.820,83 (-15,5%)	32.890,0 (-15,28%)	32.820,83 (15,5%)
2001	37.907,26	-	37.907,26 (0%)	38.431,8 (1,38%)	37.907,26 (0%)
2002	42.107,62	-	42.107,62 (0%)	42.230,0 (0,29%)	42.107,62 (0%)
2003	51.919,44	-	51.919,44 (0%)	52.017,5 (0,19%)	51.919,44 (0%)
2004	49.549,94	-	49.549,94 (0%)	49.792,7 (0,49%)	49.549,94 (0%)
2005	51.136,23	-	51.182,07 (0,1%)	52.304,6 (2,28%)	51.182,07 (0,1%)
2006	52.355,98	52.234,59 (-0,2%)	52.464,64 (0,2%)	55.027,1 (5,1%)	52.464,64 (0,2%)

Tabela 16 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - PRODUTIVIDADE (KG/HECTARE) DE SOJA NO BRASIL

Ano	Sousa, Bonjour e Figueiredo (2006)	IBGE (%)	EMBRAPA (%)	CONAB (%)	FAO (%)
1990	1.740	-	1.732,2 (-0,5%)	1.580,0 (0,01%)	1.732,2 (-0,4%)
1991	1.580	-	1.553,3 (-1,7%)	1.580,0 (0%)	1.553,3 (-1,7%)
1992	2.027	-	2.035,2 (0,4%)	2.027,0 (0%)	2.035,2 (0,4%)
1993	2.150	-	2.124,1 (-1,2%)	2.150,0 (0%)	2.124,1 (-1,2%)
1994	2.176	-	2.163,2 (-0,6%)	2.179,0 (0,14%)	2.163,2 (-0,6%)
1995	2.221	-	2.199,8 (-1,0%)	2.221,0 (0%)	2.199,8 (-1,0%)
1996	2.250	-	2.249,3 (0%)	2.175,0 (-3,33%)	2.249,3 (0%)
1997	2.298	-	2.297,7 (0%)	2.299,0 (0,04%)	2.297,7 (0%)
1998	2.353	-	2.353,3 (0%)	2.384,0 (1,32%)	2.353,3 (0%)
1999	2.372	-	2.372,4 (0%)	2.367,0 (-0,21%)	2.372,4 (0%)
2000	2.400	-	2.403,3 (0,1%)	2.414,0 (0,58%)	2.403,3 (0,1%)
2001	2.710	-	2.710,5 (0%)	2.751,0 (1,51%)	2.710,5 (0%)
2002	2.573	-	2.573,9 (0%)	2.577,0 (0,16%)	2.573,9 (0%)
2003	2.802	-	2.802,7 (0%)	2.816,0 (0,50%)	2.802,7 (0%)
2004	2.300	-	2.300,5 (0%)	2.329,0 (1,26%)	2.300,5 (0%)
2005	2.231	-	2.230,3 (0%)	2.245,0 (0,63%)	2.230,3 (0%)
2006	2.379	2379 (0%)	2.379,6 (0%)	2.419,0 (1,68%)	2.379,6 (0%)

APÊNDICE G – COMPARAÇÕES DOS RESULTADOS DE APRESENTADOS NO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - ALGODÃO EM CAROÇO

Tabela 17 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - PRODUÇÃO (MIL TONELADAS) DE ALGODÃO EM CAROÇO NO ESTADO DE MATO GROSSO

Ano	Sousa, Bonjour e Figueiredo (2006)	CONAB (%)
1990	57,634	58,3 (1,15%)
1991	73,458	107,4 (46,2%)
1992	67,862	85,5 (25,99%)
1993	85,641	96 (12,09%)
1994	91,828	122 (32,85%)
1995	87,458	111,1 (27,03%)
1996	73,553	94,7 (28,75%)
1997	78,376	99,4 (26,82%)
1998	271,038	269,2 (-0,67%)
1999	630,406	603,8 (-4,22%)
2000	1.002,836	872,3 (-13,01%)
2001	1.525,376	1.376 (-9,79%)
2002	1.141,211	1.008,8 (-11,60%)
2003	1.065,779	1.072,1 (0,59%)
2004	1.884,315	1.584,8 (-15,89%)
2005	1.682,839	1.544,5 (-8,22%)
2006	1.421,294	1.317,6 (-7,29%)

Tabela 18 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - PRODUTIVIDADE (KG/HECTARE) DE ALGODÃO EM CAROÇO NO ESTADO DE MATO GROSSO

Ano	Sousa, Bonjour e Figueiredo (2006)	CONAB(%)
1990	1.327	1.355 (2,11%)
1991	1.073	1.475 (37,46%)
1992	1.260	1.500 (19,04%)
1993	1.230	1.600 (30,08%)
1994	1.390	1.680 (20,86%)
1995	1.260	1.530 (21,42%)
1996	1.335	1.630 (22,09%)
1997	1.854	1.800 (-2,91%)
1998	2.545	2.450 (-3,73%)
1999	3.149	2.970 (-5,68%)
2000	3.890	3.250 (-16,45%)
2001	3.699	3.510 (-5,1%)
2002	3.478	3.225 (-7,27%)
2003	3.668	3.570 (-2,67%)
2004	4.011	3.615 (-9,87%)
2005	3.488	3.420 (-1,94%)
2006	3.628	3.600 (-0,77%)

Tabela 19 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - PRODUÇÃO (MIL TONELADAS) DE ALGODÃO EM CAROÇO NO BRASIL

Ano	Sousa, Bonjour e Figueiredo (2006)	CONAB(%)
1990	1.783,175	1.894,3 (6,23%)
1991	2.041,123	2.073,6 (1,59%)
1992	1.863,077	1.920,7 (3,09%)
1993	1.127,364	1.194,5 (5,95%)
1994	1.350,814	1.434,9 (6,22%)
1995	1.441,526	1.534,7 (6,46%)
1996	952,013	1.171,7 (23,07%)
1997	821,271	874 (6,42%)
1998	1.172,017	1.174,4 (0,2%)
1999	1.477,03	1.443,9 (-2,24%)
2000	2.007,102	1.887,7 (-5,94%)
2001	2.643,524	2.460,7 (6,91%)
2002	2.166,014	2.011,1 (-7,15%)
2003	2.199,268	2.212,3 (0,59%)
2004	3.798,48	3.408,6 (-10,26%)
2005	3.666,16	3.397 (-7,34%)
2006	2.882,482	2.723,59 (-5,51%)

Tabela 20 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE SOUSA, BONJOUR E FIGUEIREDO, 2006 E DA FERRAMENTA - PRODUTIVIDADE (KG/HECTARE) DE ALGODÃO EM CAROÇO NO BRASIL

Ano	Sousa, Bonjour e Figueiredo (2006)	CONAB(%)
1990	1.281	964 (-24,74%)
1991	1.373	1.056 (-23,08%)
1992	1.168	967 (-17,20%)
1993	1.221	940 (-23,01%)
1994	1.273	1.117 (-12,25%)
1995	1.306	1.249 (-4,36%)
1996	1.278	1.230 (-4,42%)
1997	1.323	1.329 (0,45%)
1998	1.420	1.335 (-5,98%)
1999	2.206	2.142 (-2,9%)
2000	2.503	2.291 (-8,46%)
2001	3.020	2.834 (-6,15%)
2002	2.848	2.690 (-5,54%)
2003	3.086	3.010 (-2,46%)
2004	3.302	3.099 (-6,14%)
2005	2.913	2.906,38 (-0,22%)
2006	3.209	3.181,02 (-0,87%)

APÊNDICE H – COMPARAÇÕES DOS RESULTADOS DE APRESENTADOS NO COSTA, 2016 E DA FERRAMENTA - SOJA

Tabela 21 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE COSTA, 2016 E DA FERRAMENTA - ÁREA PLANTADA (MIL HECTARES) DE SOJA POR REGIÃO DO BRASIL NO ANO DE 2015

Região	Costa, (2016)	IBGE (%)	CONAB (%)
Brasil	57.996,2	32.135,46 (-44,59%)	32.092,9 (-44,77%)
Centro-Oeste	22.853,7	14.669,71 (-35,81%)	14.616,1 (-36,04%)
Nordeste	8.143,6	2.867,95 (-64,78%)	2.845,3 (-65,06%)
Norte	2499,3	1.431,49 (-42,72%)	1.441,2 (-42,33%)
Sudeste	5.104,4	2.086,69 (-59,11%)	2.116,2 (-58,54%)
Sul	19.395,2	11.079,62 (-42,87%)	11.074,1 (-42,90%)

Tabela 22 – COMPARAÇÃO DE RESULTADOS DO TRABALHO DE COSTA, 2016 E DA FERRAMENTA - PRODUÇÃO (MIL TONELADAS) DE SOJA POR REGIÃO DO BRASIL NO ANO DE 2015

Região	Costa, (2016)	IBGE (%)	CONAB (%)	EMBRAPA (%)
Brasil	96.228,0	97.043,71 (0,8%)	96.228,0 (0%)	97.464,94 (1,2%)
Centro-Oeste	43.968,6	43.813,08 (-0,3%)	43.968,6 (0%)	43.943,60 (0%)
Nordeste	8.084,1	8.379,61 (3,6%)	8084,1 (0%)	8.386,41 (3,7%)
Norte	4.289,5	4.251,9 (-0,8%)	4.289,5 (0%)	4.274,63 (-0,3%)
Sudeste	5.873,5	5.753,51 (-2,0%)	5.873,5 (0%)	5.930,31 (0,9%)
Sul	34.012,3	34.845,61 (2,4%)	34.012,3 (0%)	34.845,61 (2,4%)